GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

M nucleic - nucleic search, using sw model

un on: January 9, 2003, 23:01:27 ; Search time 10638.2 Seconds
(without alignments)
11968.650 Million cell updates/sec

itle: US-09-803-126-3
erfect score: 4375
equence: 1 cgctgggactg[illegible] aaaaaaaaaaaaaaaaaaaa 4375

coring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4375	100.0	4375	6	AX253080	AX253080 Sequence
2	4352	99.5	6293	6	AX253079	AX253079 Sequence
3	2384.2	54.5	6004	9	AB058686	AB058686 Homo sapi
4	2056.4	47.0	7925	9	AF418290	AF418290 Homo sapi
5	1954.4	44.7	7515	9	AF418286	AF418286 Homo sapi
6	1511.4	34.5	4208	9	AF418293	AF418293 Homo sapi
7	1511.4	34.5	4662	9	AF418292	AF418292 Homo sapi
8	1119.4	25.6	5709	9	AF418289	AF418289 Homo sapi
9	1110.6	25.4	5272	9	AK097896	AK097896 Homo sapi
10	946.4	21.6	3780	6	AX253084	AX253084 Sequence
11	946.4	21.6	4174	6	AX253082	AX253082 Sequence
12	861	19.7	5424	9	AF418288	AF418288 Homo sapi
13	577	13.2	4802	9	AF418291	AF418291 Homo sapi
14	544.2	12.4	2099	9	AK025259	AK025259 Homo sapi
15	458	10.5	147804	10	AL645647	AL645647 Mouse DNA
c 16	458	10.5	260310	2	AL645852	AL645852 Mus muscu
17	402	9.2	1099	9	AK096869	AK096869 Homo sapi
18	371.6	8.5	4462	9	AF418287	AF418287 Homo sapi
19	311.2	7.1	11876	9	AF144094	AF144094 Homo sapi
20	291.4	6.7	11769	10	AF144095	AF144095 Mus muscu
21	286.6	6.6	5351	10	AF053130	AF053130 Mus muscu
22	262.2	6.0	180956	2	AC123410	AC123410 Rattus no
c 23	262.2	6.0	200832	2	AC130970	AC130970 Rattus no
c 24	258.8	5.9	180956	2	AC123410	AC123410 Rattus no
25	258.8	5.9	200832	2	AC130970	AC130970 Rattus no
26	195.6	4.5	170056	2	AC087643	AC087643 Homo sapi
27	195.6	4.5	178515	2	AC015551	AC015551 Homo sapi
28	184.2	4.2	217584	2	AC087749	AC087749 Homo sapi
c 29	166.2	3.8	215434	2	AC019214	AC019214 Homo sapi
c 30	102	2.3	170056	2	AC087643	AC087643 Homo sapi
c 31	102	2.3	178515	2	AC015551	AC015551 Homo sapi
32	86.2	2.0	191184	2	AC120677	AC120677 Rattus no
33	86.2	2.0	195462	2	AC129460	AC129460 Rattus no
34	84.6	1.9	88406	10	AF144093	AF144093 Mus muscu
35	84.6	1.9	261188	2	AL596386	AL596386 Mus muscu
36	81.6	1.9	125020	2	AF429315	AF429315 Homo sapi
37	76.6	1.8	95255	9	AF051976	AF051976 Homo sapi
c 38	76.6	1.8	144448	9	AC109515	AC109515 Homo sapi
c 39	76.6	1.8	164429	9	AC087164	AC087164 Homo sapi
c 40	76.6	1.8	173749	2	AC020567	AC020567 Homo sapi
41	72.6	1.7	7769	4	BTU55042	U55042 Bos taurus
42	67	1.5	6276	10	MMU249706	AJ249706 Mus muscu
43	64.6	1.5	1553	9	AK024719	AK024719 Homo sapi
44	64.6	1.5	4702	9	AF132021	AF132021 Homo sapi
45	64.6	1.5	6504	9	HSM803736	AL832428 Homo sapi

ALIGNMENTS

RESULT 1
AX253080
LOCUS AX253080 4375 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 3 from Patent WO0168866.
ACCESSION AX253080
VERSION AX253080.1 GI:15096278
KEYWORDS .
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4375)
AUTHORS Brooks,A.R., Deng,G.G. and Rubanyi,G.M.
TITLE Estrogen-regulated unconventional myosin-related protein: compositions and methods of use

```
  JOURNAL   Patent: WO 0168866-A 3 20-SEP-2001;
            SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..4375
                     /organism="Mus sp."
                     /db_xref="taxon:10090"
                     /note="Mouse myosin related protein (MRP) variant 2 cDNA"
BASE COUNT       [illegible] a   1269 c   1016 g    825 t
ORIGIN

  Query Match             100.0%;  Score 4375;  DB 6;  Length 4375;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches [illegible];  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCTGGGACTGTCAGCTACCAGGTGCACAAGTTCATAAACAGAAACAGGGGCCACCTGGA  60
Db       1 CGCTGGGACTGTCAGCTACCAGGTGCACAAGTTCATAAACAGAAACAGGGGCCACCTGGA  60

Qy      61 [illegible] 120
Db      61 GCTCGCTGTGCTGGAGATGCTCAGGCAGAGCCAGCTGCAGGTGACCTAGCCTTCCTTTCA 120

Qy     121 GCTCATGGGCAGCCTGTTCCAAGAAGCAGAGCCCCAGGCTGGGACTGAGCAAAACAAACCC 180
Db     121 GCTCATGGGCAGCCTGTTCCAAGAAGCAGAGCCCCAGGCTGGGACTGAGCAAAACAAACCC 180

Qy     181 CACATTGGCCTCTCGATTCCAGCAGACCCTGGGTGACTTGCTAGCTCGGCTAGGCAGCAG 240
Db     181 CACATTGGCCTCTCGATTCCAGCAGACCCTGGGTGACTTGCTAGCTCGGCTAGGCAGCAG 240

Qy     241 GGGCCATGTCTACGTCATCCACTGTCTCAATCCCAGCCCCTGGAAAGATCCCAGGCCTCTT 300
Db     241 GGGCCATGTCTACGTCATCCACTGTCTCAATCCCAGCCCCTGGAAAGATCCCAGGCCTCTT 300

Qy     301 GTACGTGGGGCATGTGGCAGAGCAGCTGCGTCAGGCTGGCATCCTGGAGATCATAGGCAC 360
Db     301 GTACGTGGGGCATGTGGCAGAGCAGCTGCGTCAGGCTGGCATCCTGGAGATCATAGGCAC 360

Qy     361 [illegible] 420
Db     361 CCCCACTACCCACTTCCCCGTGCGAGTGTCCTTCCAAGTCTTTCTGGCAAGGTTCCATGC 420

Qy     421 [illegible] 480
Db     421 [illegible] 480

Qy     481 TAAATTCCTGGGGGCAGAGTCACCGCTGTATCATCTTGGAGTCACCCAGGTCCTGCTGCA 540
Db     481 TAAATTCCTGGGGGCAGAGTCACCGCTGTATCATCTTGGAGTCACCCAGGTCCTGCTGCA 540

Qy     541 [illegible] 600
Db     541 GGAATAGGGCTGGCAGCAGCTAGAACAGCTGTGGGCTCAGCGGCGCTCACAGGCCCTGCT 600

Qy     601 [illegible] 660
Db     601 [illegible] 660

Qy     661 GATGTAGGCTGGTGTGCGTGGGCTCCAGGCCAGGAAGCGATATCTCCAGCGGAGGTCAGC 720
Db     661 GATGTAGGCTGGTGTGCGTGGGCTCCAGGCCAGGAAGCGATATCTCCAGCGGAGGTCAGC 720

Qy     721 [illegible] 780
Db     721 TTTGGACAGCTGAACACCATTCTCCTAGTGGCCCGGCCCCTGCTCCGGAGACGACAGAA 780

Qy     781 GTTACGCTGTGCCCTGGCCCGCACAGCGGGGAGCCCTGGGGGAAAGTGTCAAATATGGA 840
Db     781 GTTATGCTGTGCCCTGGCCCGCACAGCGGGGAGCCCTGGGGGAAAGTGTCAAATATGGA 840

Qy     841 [illegible] 900
Db     841 [illegible] 900
```

```
Qy     901 CCAGGCCTTGCTGACGGGGAGCATCATAGAGTCCTGCTACTACAGCCCCTGCT 954
Db     901 CCAGGCCTTGCTGACGGGGAGCATCATAGAGTCCTGCTACTACAGCCCCTGCT 954

Qy     961 GAGCCTGACTCTCCCTCAGAGATTGACCAGTTTCCTCTCTACTTTTCTATCTATCT 1020
Db     961 GAGCCTGACTCTCCCTCAGAGATTGACCAGTTTCCTCTCTACTTTTCTATCTATCT 1020

Qy    1021 CTTTCAGAAGCCATTCTGCCTCGACTAGGGTAGTACTGAATAGCTGACTTTTT 1076
Db    1021 CTTTCAGAAGCCATTCTGCCTCGACTAGGGTAGTACTGAATAGCTGACTTTTT 1076

Qy    1077 AGATGGCGAGAACCCTCAGCAGGCTCTGGAGATCAACAGTTATCTTCTGCTCTTTT 1136
Db    1077 AGATGGCGAGAACCCTCAGCAGGCTCTGGAGATCAACAGTTATCTTCTGCTCTTTT 1136

Qy    1137 GGAAGGATCTCTGCAGTCCTGGCAAGAGCAGAGCCATGGGCATTTCTTCTCTATAAT 1196
Db    1137 GGAAGGATCTCTGCAGTCCTGGCAAGAGCAGAGCCATGGGCATTTCTTCTCTATAAT 1196

Qy    1197 CCAGCGACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGTTTTTTTTTTAGTAAT 1256
Db    1197 CCAGCGACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGTTTTTTTTTTAGTAAT 1256

Qy    1257 CAACCCAGATGAGCAACAGAATCAGTGTGGCTGGCCCTAATTTATTTCTTTATTT 1316
Db    1257 CAACCCAGATGAGCAACAGAATCAGTGTGGCTGGCCCTAATTTATTTCTTTATTT 1316

Qy    1317 CTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTTAAATTTTTATTTGACTAGTTT 1376
Db    1317 CTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTTAAATTTTTATTTGACTAGTTT 1376

Qy    1377 CAGTGGCATGGCAGCCCTGTGCCAGCACAAGCTGTTAGGTTTTTGAGCAGATAGTTT 1436
Db    1377 CAGTGGCATGGCAGCCCTGTGCCAGCACAAGCTGTTAGGTTTTTGAGCAGATAGTTT 1436

Qy    1437 GGCTCCCATGGCTTGGAGGTCCCACCCACCCACAAATTTGAAAAAATGTTTTTAATT 1496
Db    1437 GGCTCCCATGGCTTGGAGGTCCCACCCACCCACAAATTTGAAAAAATGTTTTTAATT 1496

Qy    1497 TCGGGGCCGCATGGCGCTGGATGTGTTCACATTCAAAGAGAAAGCTAATTTTTGAA 1556
Db    1497 TCGGGGCCGCATGGCGCTGGATGTGTTCACATTCAAAGAGAAAGCTAATTTTTGAA 1556

Qy    1557 GGAATCCTGGACCACGGGAGAGCAGTTTGCAGGGTGATTTAAAGAGAGAGTTTTGA 1616
Db    1557 GGAATCCTGGACCACGGGAGAGCAGTTTGCAGGGTGATTTAAAGAGAGAGTTTTGA 1616

Qy    1617 GGCGCCCCCTCGTGGCTGGTCTGTGTCACTCAATCTGGGAATTTGGT 1676
Db    1617 GGCGCCCCCTCGTGGCTGGTCTGTGTCACTCAATCTGGGAATTTGGT 1676

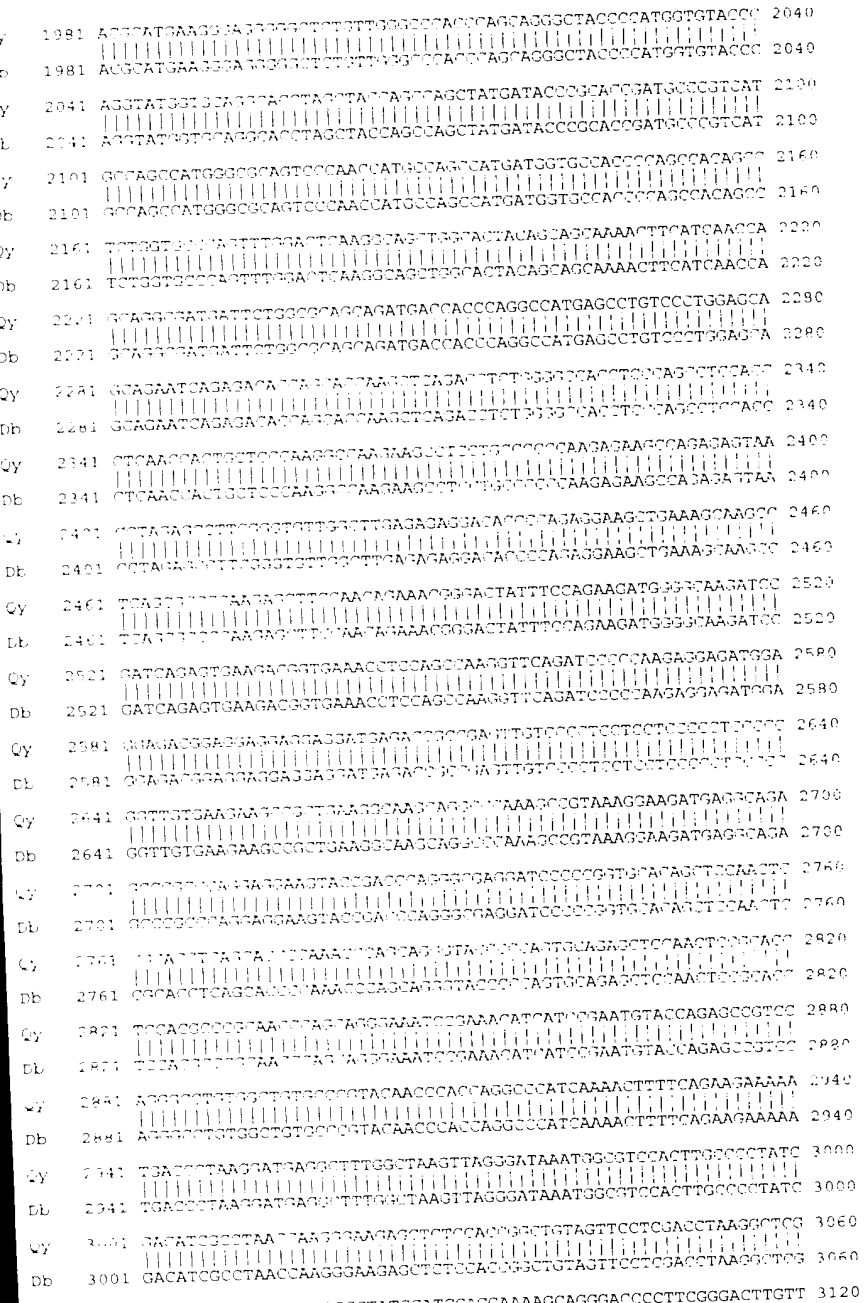
Qy    1677 TGGCTGTGACTTTGTGTTGGACCTAATAGGCCAGACTGAGGATTTGGAGACCCAGTTG 1736
Db    1677 TGGCTGTGACTTTGTGTTGGACCTAATAGGCCAGACTGAGGATTTGGAGACCCAGTTG 1736

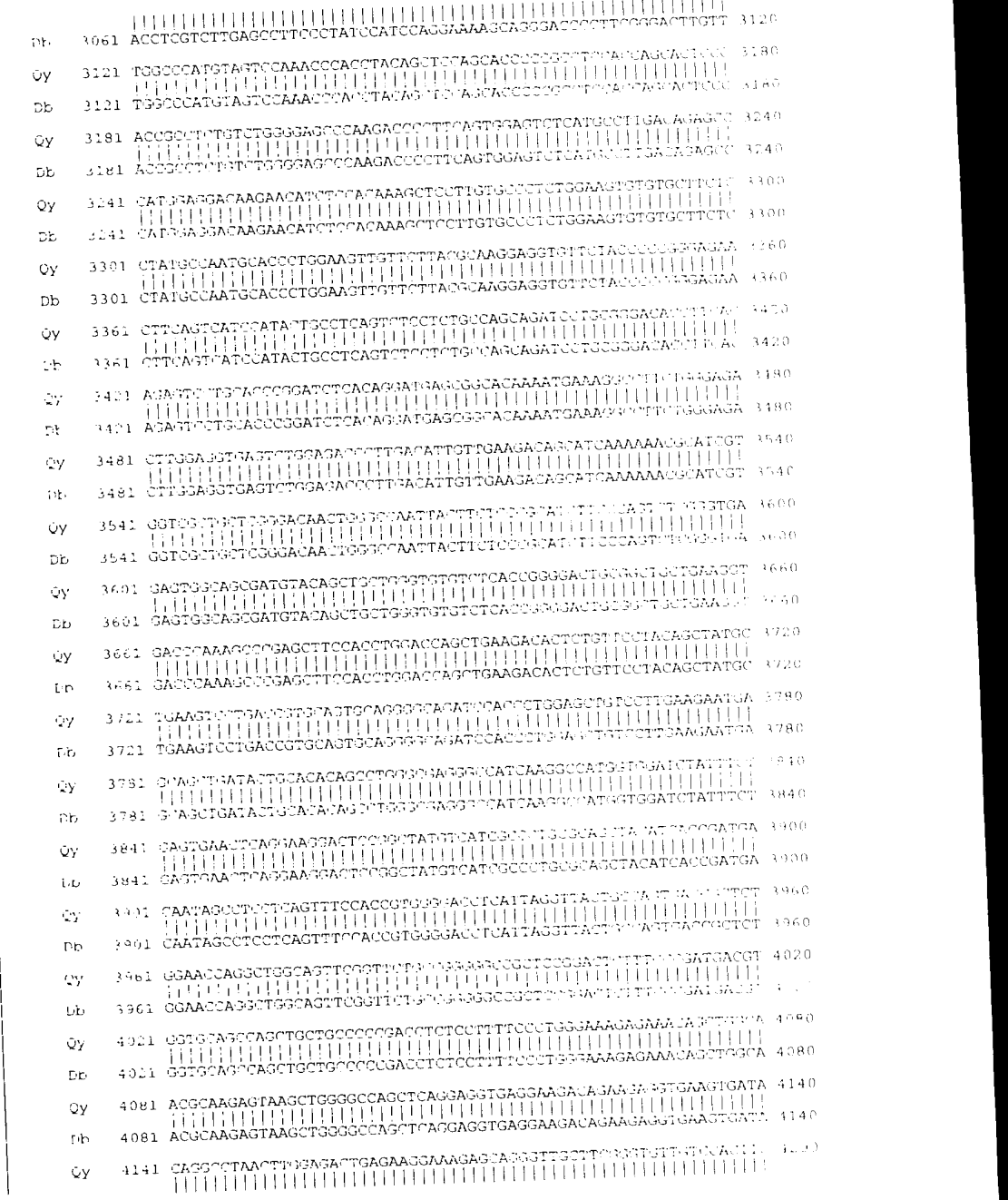
Qy    1737 TCCCCACAACTACCCCATCACTCCTCTTGGTTTAGCTGAGAGTATCCTCCAGCCCTGG 1796
Db    1737 TCCCCACAACTACCCCATCACTCCTCTTGGTTTAGCTGAGAGTATCCTCCAGCCCTGG 1796

Qy    1797 TGTCCAGGCTCCTTCCCTGCCCCCAGGACTCCTCCAGGTCCAGCCCCAATACTGGCCAG 1856
Db    1797 TGTCCAGGCTCCTTCCCTGCCCCCAGGACTCCTCCAGGTCCAGCCCCAATACTGGCCAG 1856

Qy    1857 CAGCCGCCCTCCGGGCGAGGCCAGTAAGCCTGACAACCTGGATCCTTTCCTTGACTCACCT 1916
Db    1857 CAGCCGCCCTCCGGGCGAGGCCAGTAAGCCTGACAACCTGGATCCTTTCCTTGACTCACCT 1916

Qy    1917 CTTTGAACCAGCGCCTCGCTCCGGGTTTCAGTGATCTGGAAGAAGCCTCCCCTGAGTAT 1976
Db    1917 CTTTGAACCAGCGCCTCGCTCCGGGTTTCAGTGATCTGGAAGAAGCCTCCCCTGAGTAT 1976
```





```
Db    4141 'AGG CTAAC TTGGAGAC TGAGAAGGAAAGAGCAGGGTTGCTTCGGGTGTTGTCCACTTC 4200

Qy    4261 TCTCTCTCCTGGCCAGGGCTCAATGTGTTCCTGTCCTTTACCATCTCCTGACTTTTTGCC 4260

Db    4201 CTGTCCTGGTGGCCAGGGCTCAATGTGTTCCTGTCCTTTACCATCTCCTGACTTTTTGCC 4260

Qy    4261 ATTTGTGAGACTCTAAATCTACACCCTCTAACTCTGGTACTTAGTTCAGTGTCTCCATAGA 4320

Db    4261 ATTTGTGAGACTCTAAATCTACACCCTCTAACTCTGGTACTTAGTTCAGTGTCTCCATAGA 4320

Qy    4321 GGATGCTTAATAAATAACCTTGGTTTTCCTGGAAAAAAAAAAAAAAAAAAAAAAAA 4375

Db    4321 GGATGCTTAATAAATAACCTTGGTTTTCCTGGAAAAAAAAAAAAAAAAAAAAAAAA 4375


RESULT 2
AX253079
LOCUS       AX253079                6293 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 2 from Patent WO0168866.
ACCESSION   AX253079
VERSION     AX253079.1  GI:15986277
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6293)
  AUTHORS   Brooks,A.R., Deng,G.G. and Rubanyi,G.M.
  TITLE     Estrogen-regulated unconventional myosin-related protein:
            Compositions and methods of use
  JOURNAL   Patent: WO 0168866-A 2 20-SEP-2001;
            SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..6293
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
                     /note="Mouse myosin related protein (MRP) variant 1 cDNA"
BASE COUNT     1433 a   1917 c   1733 g   1210 t
ORIGIN

  Query Match             99.5%;  Score 4352;  DB 6;  Length 6293;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4352;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TGTTTTTAATTTCAGTTATCAGGTGCAGAAGTTCATAAACAGAAACAGGGGCCACCTGGA 60

Db       1 TGTTTTGGATTGTCACTTACAGGTGCAGAAGTTCATAAACAGAAACAGGGGCCACCTGGA 60

Qy      61 GCTTCTGTTGCTGGAGATGCTCAGGCAGAGGCAGCTGCAGGTGACCTAGCCTTCCTTTCA 120

Db      61 GCTTCTGTTGCTGGAGATGTTCAGGCAGAGGCAGCTGCAGGTGACCTAGCCTTCCTTTCA 120

Qy     121 ACTTATGGGCAGCCTGTTCCAAGAAGCAGAGCCCCAGGCTGGGACTGAGCAAAACAAACC 180

Db     121 ACTTATGGGCAGCCTGTTCCAAGAAGCAGAGCCCCAGGCTGGGACTGAGCAAAACAAACC 180

Qy     181 CATATTGGGTCTCTGGATTCCAGCAGACCCTGGGTGACTTGCTAGCTCGGCTAGGCAGCAG 240

Db     181 CATATTGGGTCTCTGGATTCCAGCAGACCCTGGGTGACTTGCTAGCTCGGCTAGGCAGCAG 240

Qy     241 GGTCCATGTCTACGTCATCCACTGTCTCAATCCCACCCCTGGAAAGATCCCAGGCCTCTT 300

Db     241 GGTCCATGTCTACGTCATCCACTGTCTCAATCCCACCCCTGGAAAGATCCCAGGCCTCTT 300

Qy     301 GGACTTGGGGCATGTGGCAGAGCAGCTGCGTCAGGCTGGCATCCTGGAGATCATAGGCAC 360

Db     301 GGACTTGGGGCATGTGGCAGAGCAGCTGCGTCAGGCTGGCATCCTGGAGATCATAGGCAC 360

Qy     361 CTGAGTACTCCACTTCCCCGTGCTGAGTGTCCTTCCAAGTCTTTCTGGCAAGGTTCCATGC 420

Db     361 CTGGAGTACTCCACTTCCCCGTGCTGAGTGTCCTTCCAAGTCTTTCTGGCAAGGTTCCATGC 420

Qy     421 CCTGGGGTCAGGGAGACAGAAAGCTGCCTCTGACCAGGAGAGGTGTGGTGCCATCCTCAG 480

Db     421 CCTGGGGTCAGGGAGACAGAAAGCTGCCTCTGACCAGGAGAGGTGTGGTGCCATCCTCAG 480

Qy     481 TGAAGTGCTGGGGGCAGAGTTACCGTTGTATCATTTTGAATTACTAGTCTTTAGTTTA 540

Db     481 TGAAGTGCTGGGGGCAGAGTTACCGTTGTATCATTTTGAATTACTAGTCTTTAGTTTA 540

Qy     541 GGAACAGGGCTGGCAGCAGCTAGAACAGCTGTGGGTTACTAGGCTTTGTT 590

Db     541 GGAACAGGGCTGGCAGCAGCTAGAACAGCTGTGGGTTACTAGGCTTTGTT 590

Qy     591 CACTCTGCACCGTGGCTCCGAGCCTGTATCACCGGAAGTCTCTTTTTTTTT 640

Db     591 CACTCTGCACCGTGGCTCCGAGCCTGTATCACCGGAAGTCTCTTTTTTTTT 640

Qy     641 CATGCAGGCTCGTGTGCCTGGCTCCAGGGCAGGAAGTGATACTCAAGAGTTAT 690

Db     641 CATGCAGGCTCGTGTGCCTGGCTCCAGGGCAGGAAGTGATACTCAAGAGTTAT 690

Qy     691 TCTGGCACAGCTGAACACCATTCTCCTAGTGGCTGGCTCTTTGTGGAGACAGAAA 740

Db     691 TCTGGGACAGCTGAACACCATTCTCCTAGTGGCTGGCTCTTTGTGGAGACAGAAA 740

Qy     741 GCTACGGTGTGCCCCTGGCTCGCACAGTGGGGAGCTCTTGTTGAAATCTTAAATATTGA 800

Db     741 GCTACGGTGTGCCCCTGGCTCGCACAGTGGGGAGCTCTTGTTGAAATCTTAAATATTGA 800

Qy     801 CCTGGGTCTGCTTAGAGATCCCCTGGCCAGTGGCTACTCTTGTTGAGATTTTGGAAGGTA 860

Db     801 CCTGGGTCTGCTTAGAGATCCCCTGGCCAGTGGCTACTCTTGTTGAGATTTTGGAAGGTA 860

Qy     861 CCAGGCCTTGCTGACGGGGAGCACTACAGAGTCCTCAGTTCTAGTTTTGTATCCACTAG 920

Db     861 CCAGGCCTTGCTGACGGGGAGCACTACAGAGTCCTCAGTTCTAGTTTTGTATCCACTAG 920

Qy     921 CAGCCTGACTCTCCCTCCAGACATGACCAGTTTTCTTCTACTTTTTATCCACTAG 980

Db     921 CAGCCTGACTCTCCCTCCAGACATGACCAGTTTTCTTCTACTTTTTATCCACTAG 980

Qy     981 CTTTCAGAAGCCATTTCTGCCTGGACCAGGGAGCCACTGGACTAGTCTTTCACTGGTT 1040

Db     981 CTTTCAGAAGCCATTTCTGCCTGGACCAGGGAGCCACTGGACTAGTCTTTCACTGGTT 1040

Qy    1041 AGATGGCGAGAACCCTCAGCAGGCTCTGGAGATCAACAGGGAACTCTTGGCTCTGGG 1100

Db    1041 AGATGGCGAGAACCCTCAGCAGGCTCTGGAGATCAACAGGGAACTCTTGGCTCTGGG 1100

Qy    1101 GGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACATGGGCATTTTCTTCTGCAGTAGT 1160

Db    1101 GGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACATGGGCATTTTCTTCTGCAGTAGT 1160

Qy    1161 CCAGCGACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGTTTCTCAGTGTGGGG 1220

Db    1161 CCAGCGACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGTTTCTCAGTGTGGGG 1220

Qy    1221 CAACCCAGATGAGCAACAGAATCAGCGTGGCTGGGCCCTAATGTTATTCTCTCTA 1280

Db    1221 CAACCCAGATGAGCAACAGAATCAGCGTGGCTGGGCCCTAATGTTATTCTCTCTA 1280

Qy    1281 CTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTTTATCTGACAGGTCC 1340

Db    1281 CTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTTTATCTGACAGGTCC 1340

Qy    1341 CAGTGGCATGGCAGCCCTGTGCCAGCACAAGCTGTTAGGTGCCCTGAGCACAGACCCCT 1400

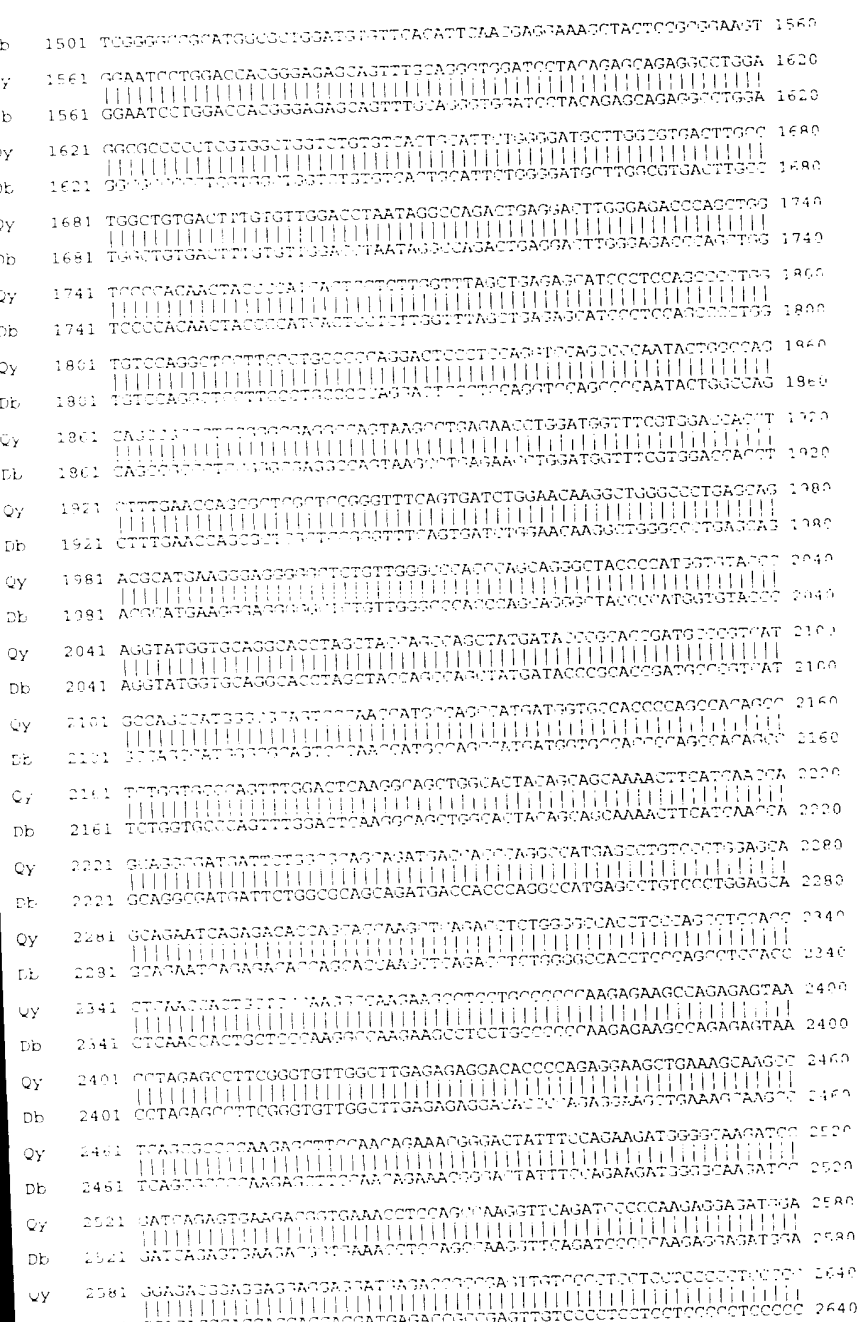
Db    1341 CAGTGGCATGGCAGCCCTGTGCCAGCACAAGCTGTTAGGTGCCCTGAGCACAGACCCCT 1400

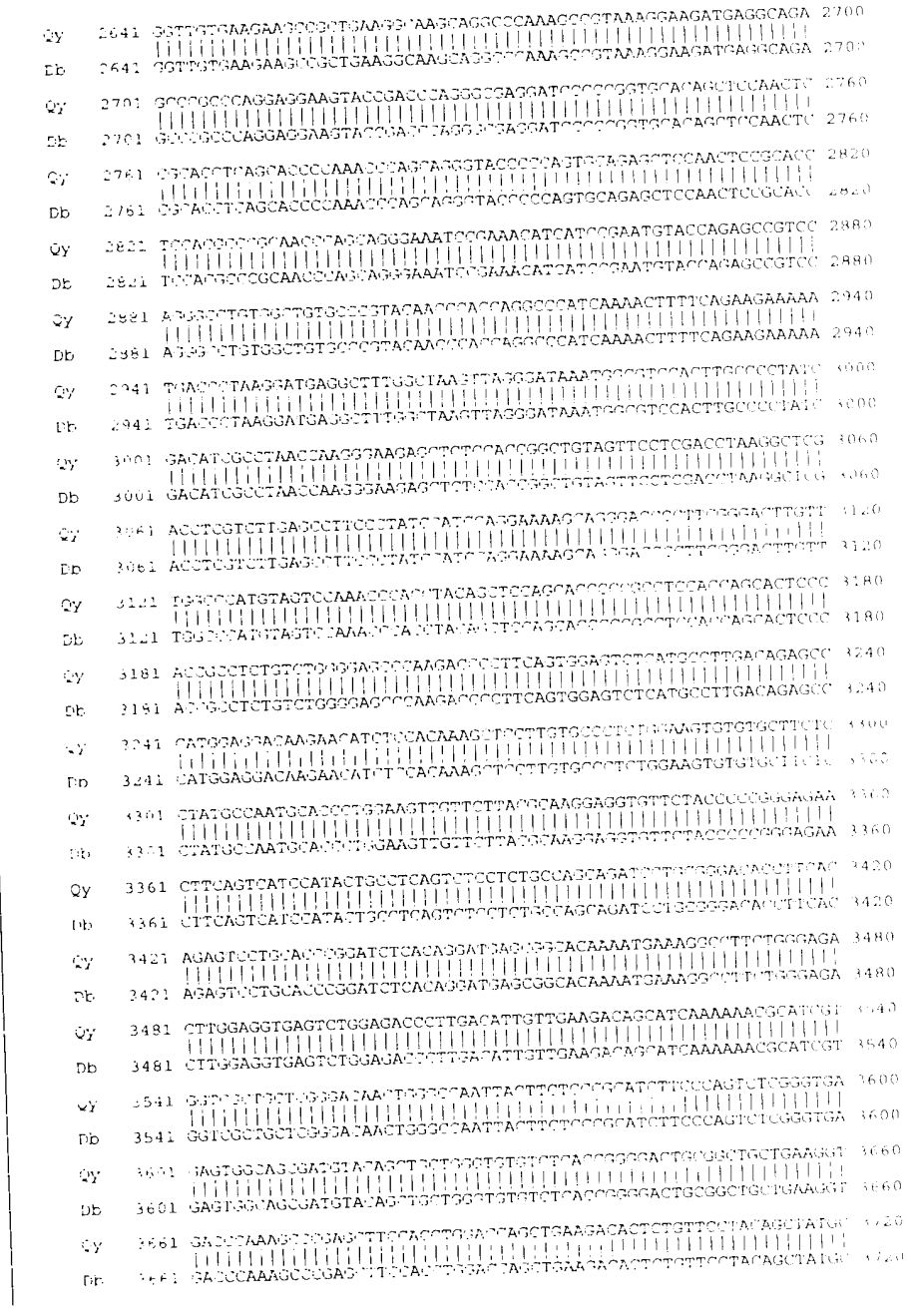
Qy    1401 GGCTCCCATGGCTTCGAGGTCCCACCCACCCACAACTTGAATCGAAGTTGGTTTTACG 1460

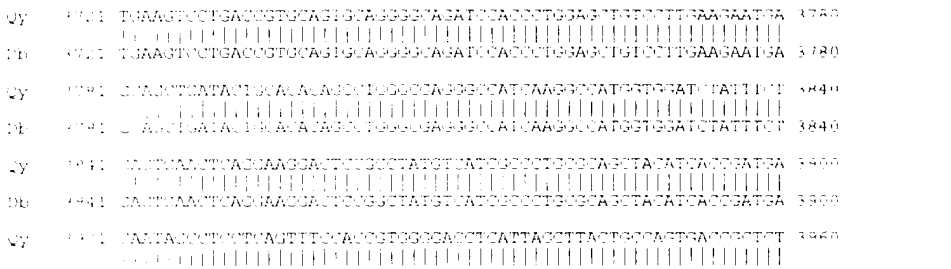
Db    1401 GGCTCCCATGGCTTCGAGGTCCCACCCACCCACAACTTGAATCGAAGTTGGTTTTACG 1460

Qy    1461 TCGGGGCCCCACTGGGGCTGGAGTCTGTCAATTCAAGCAGAAAGTACTTTGGGAAGT 1520

Db    1461 TCGGGGCCCCACTGGGGCTGGAGTCTGTCAATTCAAGCAGAAAGTACTTTGGGAAGT 1520
```



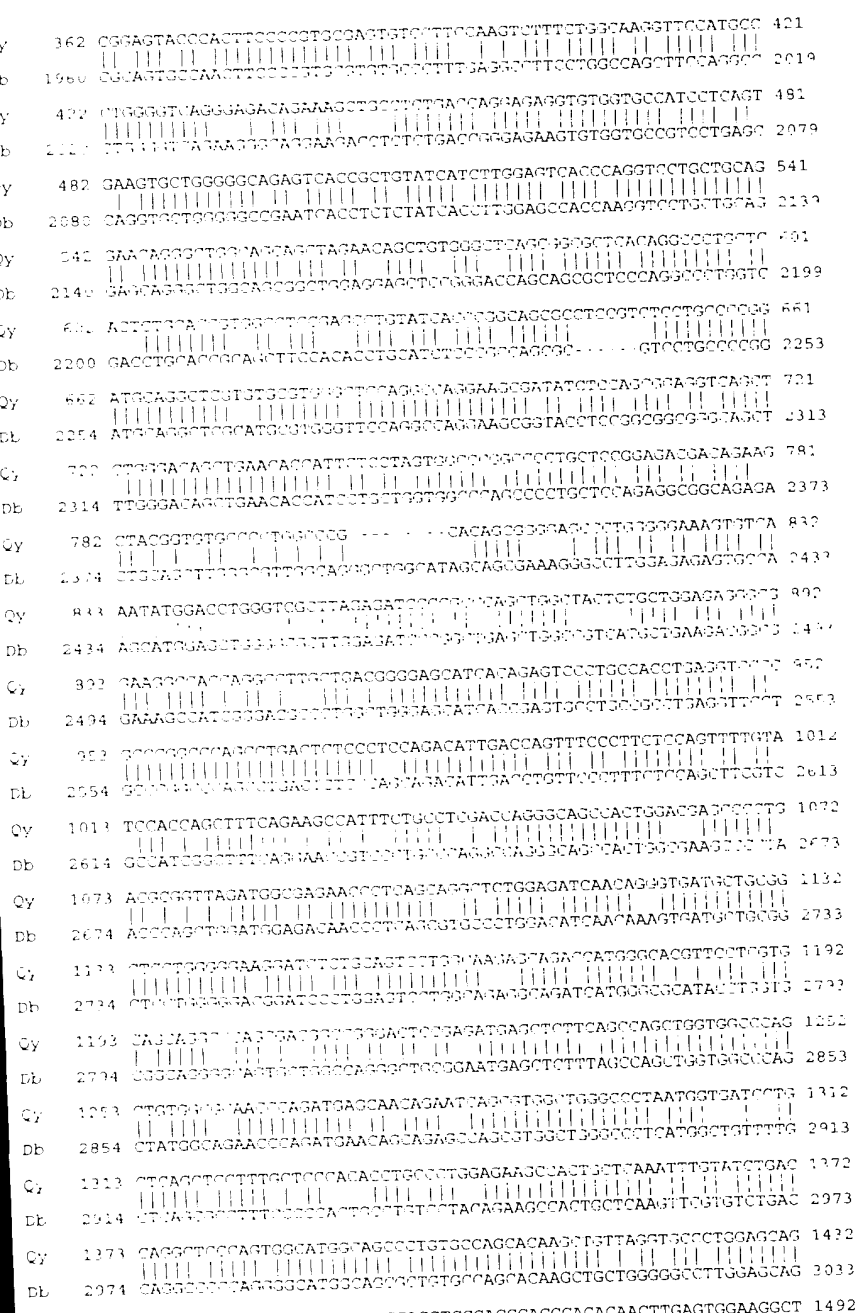


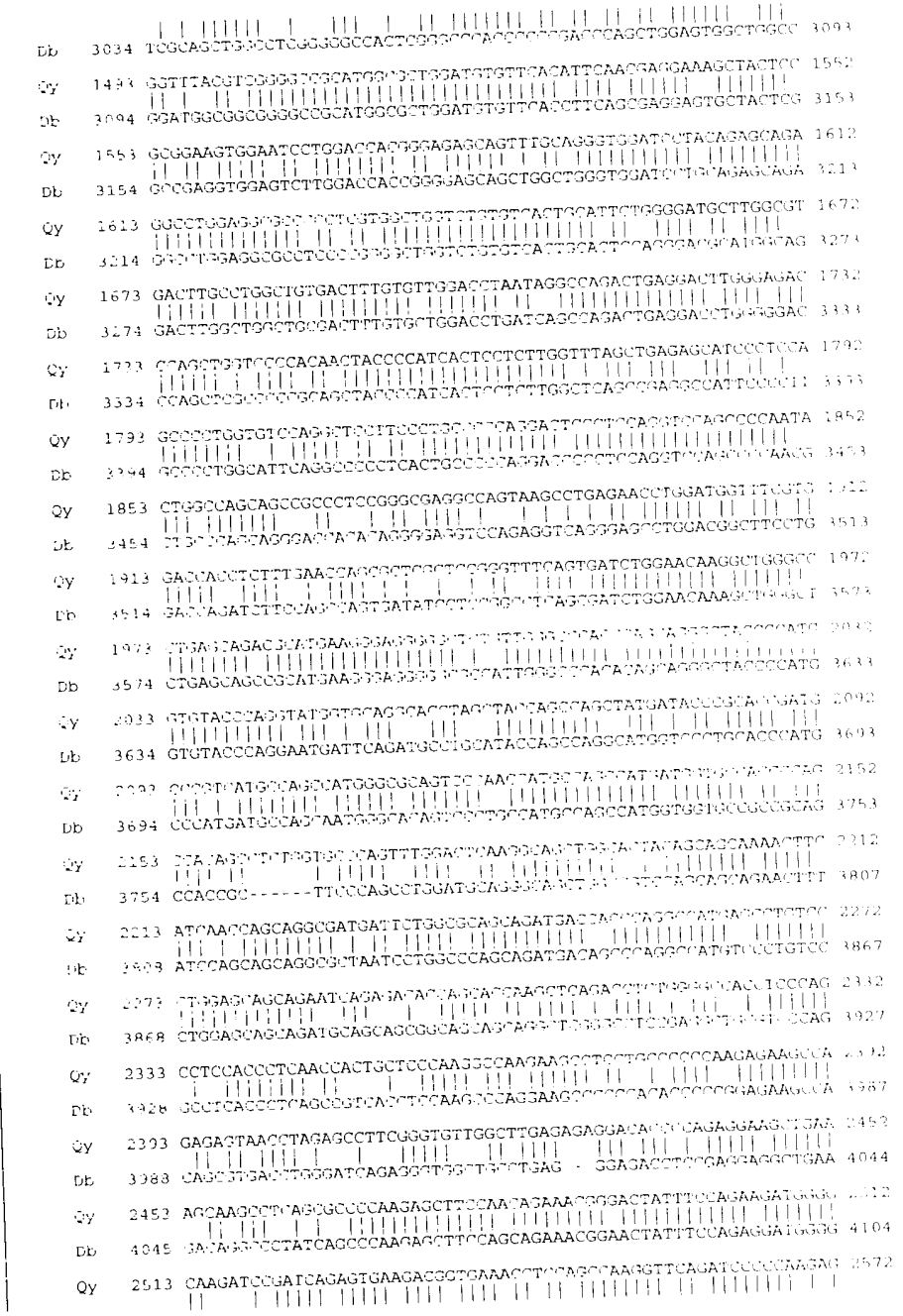


```
RESULT 1
AB058686
LOCUS       AB058686               6004 bp    mRNA    linear   PRI 05-JUN-2001
DEFINITION  Homo sapiens mRNA for KIAA1783 protein, partial cds.
ACCESSION   AB058686
VERSION     AB058686.1  GI:14017782
KEYWORDS    .
SOURCE      Homo sapiens brain cDNA to mRNA, clone:fg01285.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. XX.
            The complete sequences of 100 new cDNA clones from brain which code
            for large proteins in vitro
  JOURNAL   DNA Res. 8 (2), 85-95 (2001)
  MEDLINE   21245130
REFERENCE   2  (bases 1 to 6004)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge, Tel:81 438 52 3913,
            Fax:81 438 52-3914)
FEATURES             Location/Qualifiers
     source          1..6004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="fg01285"
                     /tissue_type="brain"
                     /note="vector:pBluescriptII XK plus"
     gene            1..6004
                     /gene="KIAA1783"
     CDS             <1081..5763
                     /gene="KIAA1783"
                     /note="Start codon is not identified."
                     /codon_start=1
                     /product="KIAA1783 protein"
                     /protein_id="BAB47412.1"
                     /db_xref="GI:14017783"
                     /translation="[illegible]
                     CPPHCHMPPGPAGEWPGATVQQPRQRAPTALLQTIAAGPGGGVMSAGVAVLSAGTGA
                     [illegible]
                     GDHPSYAKPRLPLPVFTVRHYAGTVTYQVHEFLNRNKDPLDIAVVEMSKCSKLVSE
                     [illegible]
                     HVTEQLHQAAILEAVGTRSANFTVRVIPPAPLASPLALSHGQLEDSGREFTAAVIL
                     [illegible]
                     QARMRGPQAPKRYLRRRAALGQLNTILLVALTLQPRQKLQLGRWQEWHVTPEAIRNT
                     [illegible]
                     SFVAIGFQEPSLPPPQQPLAKPLTQLDGDNPQPAIDINEVMLELLGPGCLEEWQPLIM
                     GAYLVPDGQTWPGLRNPLLFRQLVAQIWQNTLPQLQLRGWAIMAVLLSAFPPLLPVLGT
                     LLKFVSIGARRQMAADQHHILQAIRQGLASQTATRAHTPTGLEWLAGWPQSEMAILV
                     [illegible]
                     DLISQTEDLGDPARPRSYPITPLGSAEAIPLAPTTQAISLPPGPTTSGPAPTLSGPHT
                     [illegible]
                     IQMPAYQPGMVPAPMPMMPAMGTVPAMPAMVVIPLPTLPGLIAVGLQLNPLLLLA
                     [illegible]
                     LGSEGGCLRETSEEAELRPYQPKSPQQKRNYFQRMGLPQITVFTMPTTAKVDIDLGHA
                     [illegible]
                     GHGQGPAQGPGIVVRSSDSKPPRPQPSRERDRAIIRMSQSRPQPVPVPVQLCPEPRPAP
                     [illegible]
                     PLLDLFGQKLPIAHTPPPPPAPPLPLPEDPSTLSAFRPSTQQVPPQGVVSTQLLAIGS
                     [illegible]
                     MKDLLGGLEVDLDSLTTTEDSVKKFIVVAAKINWAYFQRFFIVSGRGGQIVQLLAVL
                     [illegible]
                     RAIEALVELFLNELKKDSGYVIALPQSYIQN[illegible]AITLRLSWLPI
                     [illegible]
                     SAGGRSGLFPADIVQEAAARDFSFGREQAISWRKTGLQNRDPQAPWIPAAFEVEFMGP
                     GQAEARPA"
BASE COUNT     1134 a   1942 c   1870 g   1058 t
ORIGIN
```



```
  Query Match             94.5%;  Score 8394.2;  DB 4;  Length 6004;
  Best Local Similarity   75.7%;  Pred. No. 0;
  Matches 2007;  Conservative    0;  Mismatches  645;  Indels   66;  Gaps    4;
```

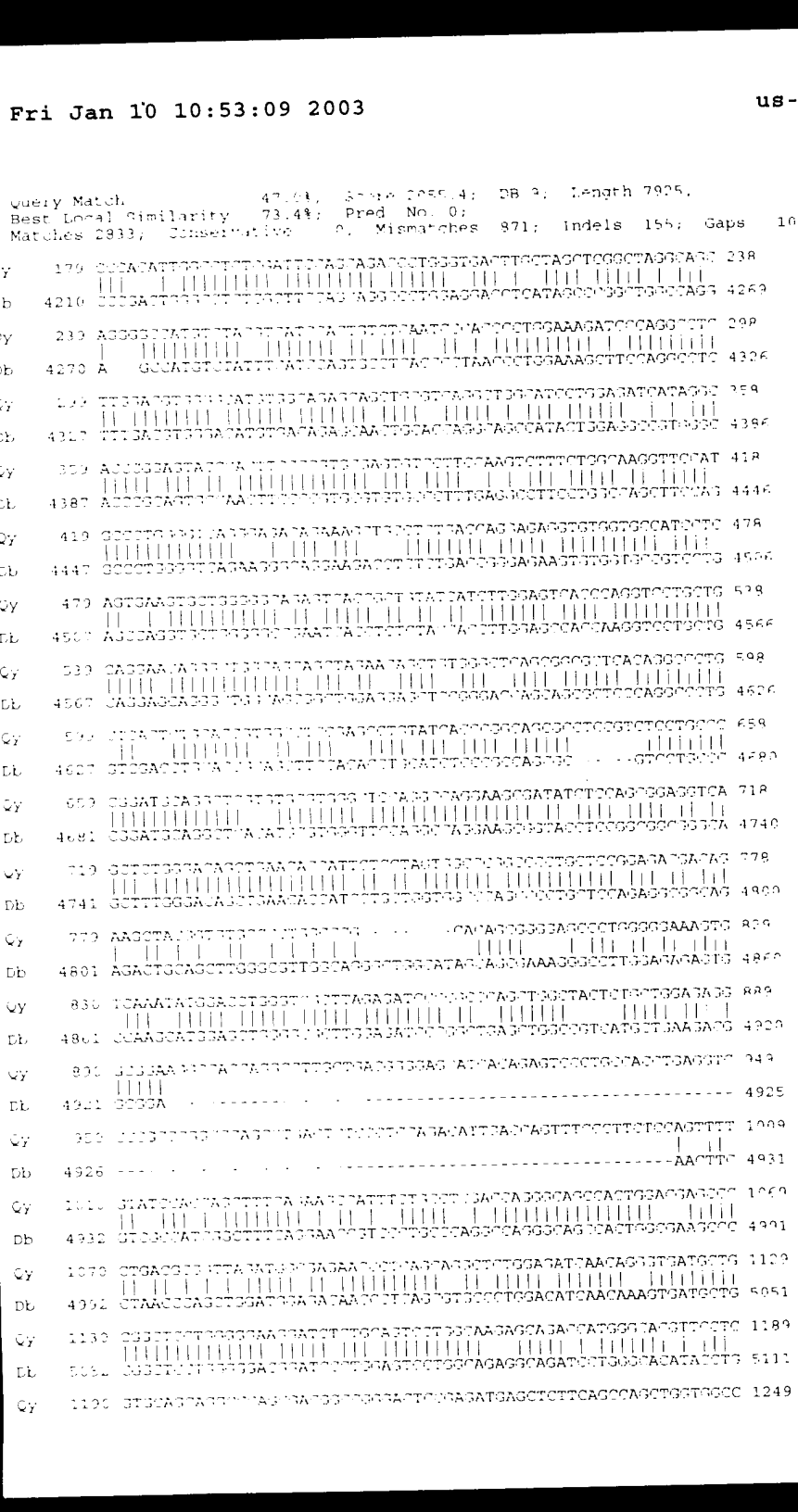


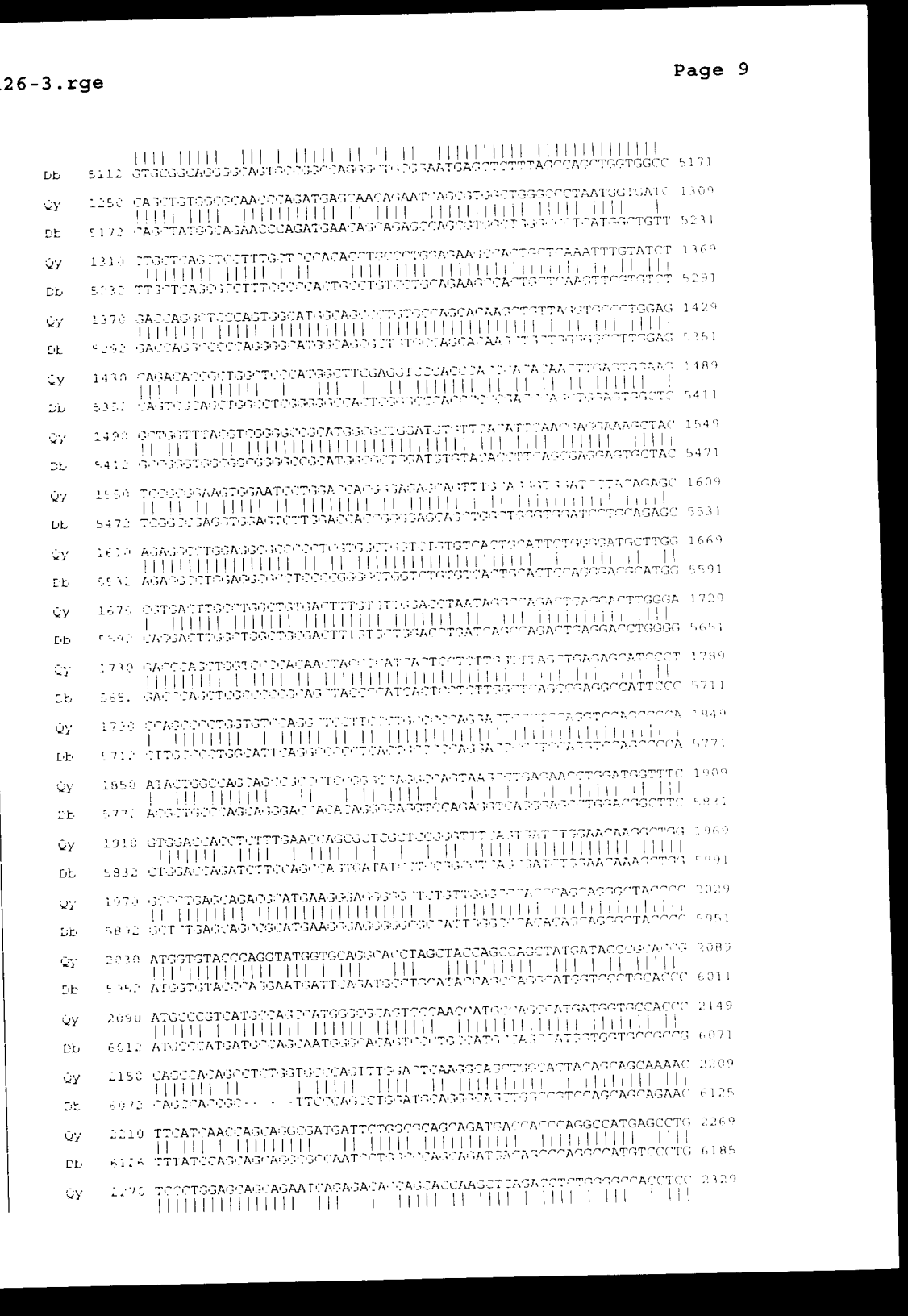


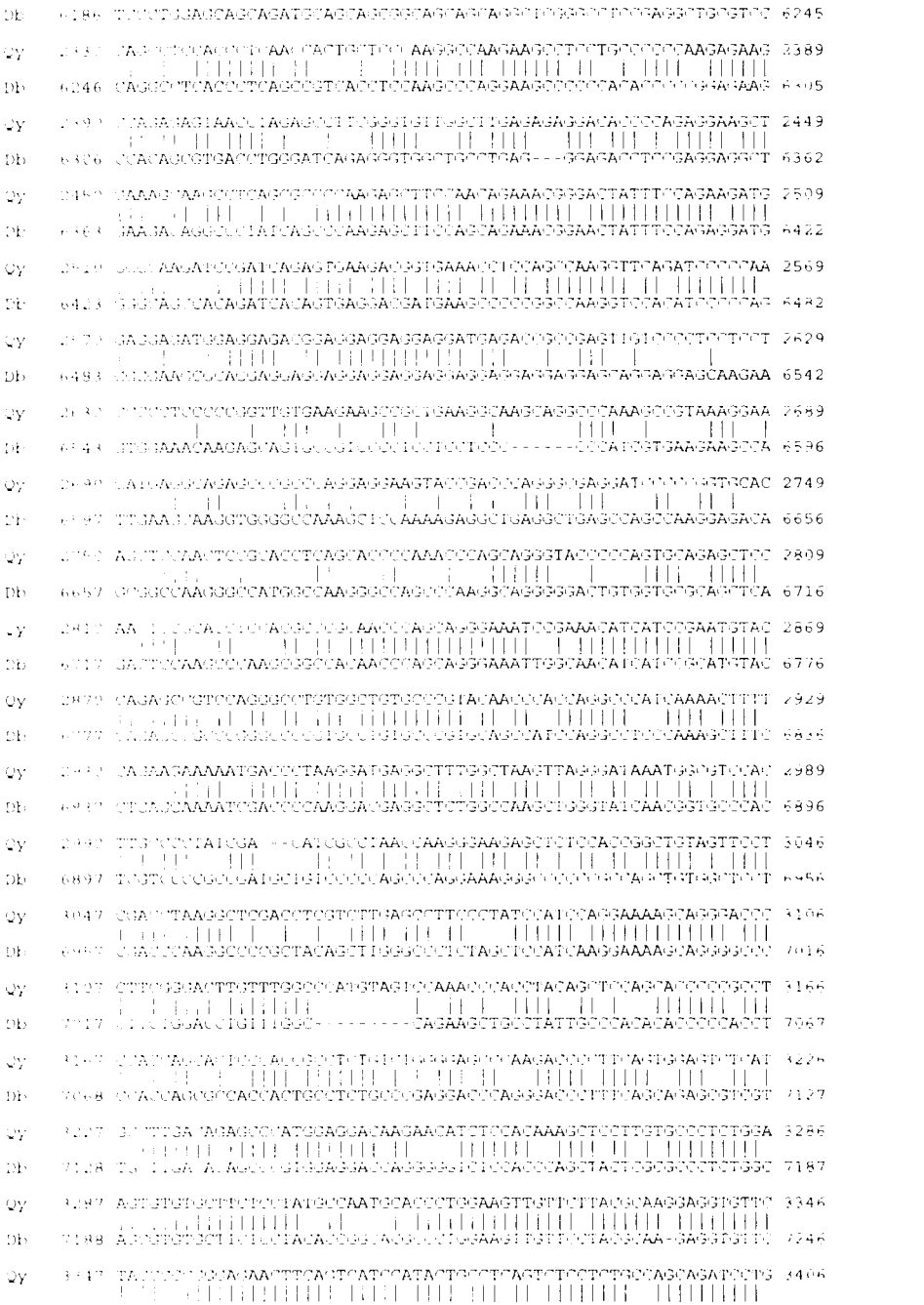


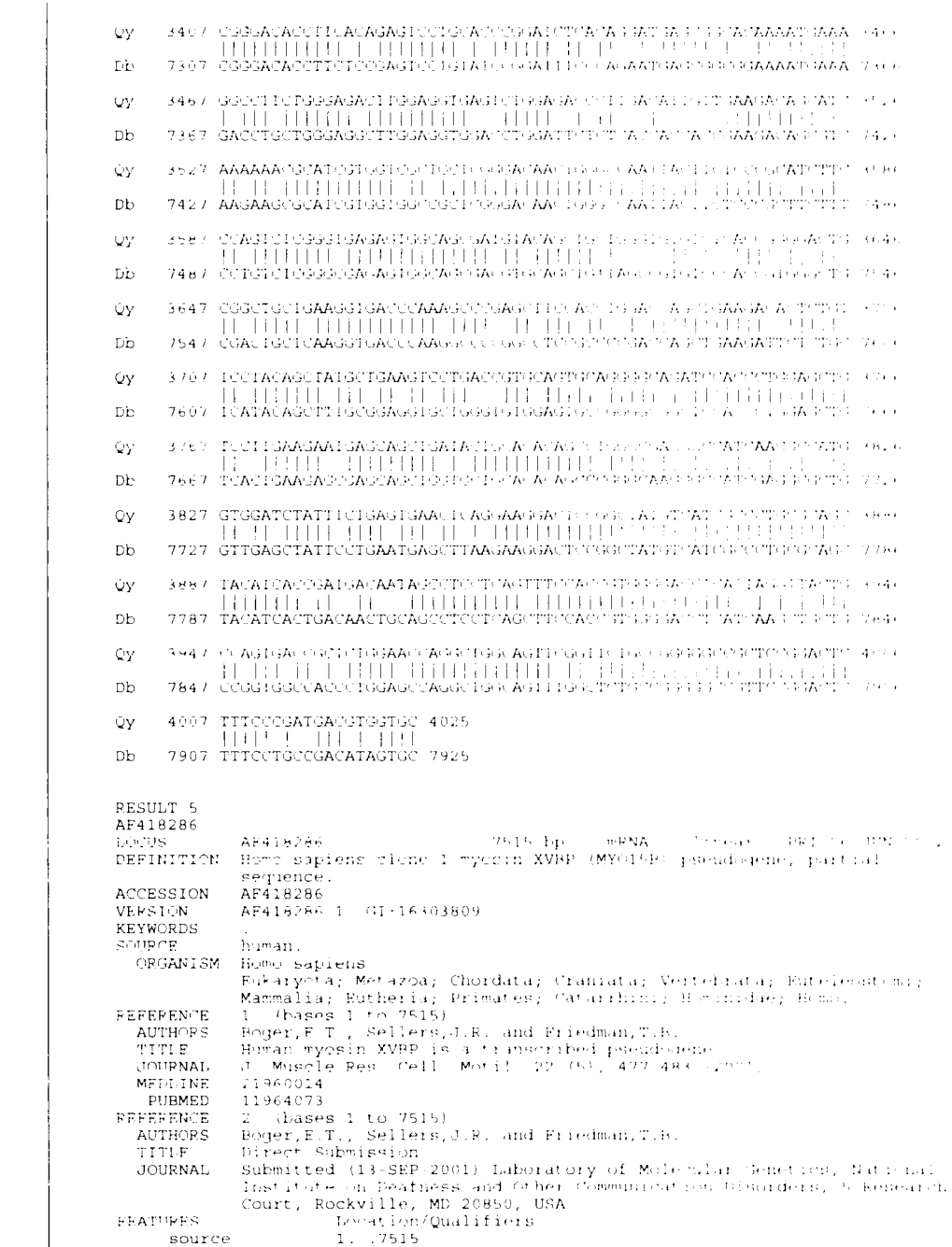
```
RESULT 4
AF418290
LOCUS       AF418290                7925 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens clone 15 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418290
VERSION     AF418290.1  GI:16903813
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7925)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-485 (2001)
  MEDLINE   21960014
   PUBMED   11964073
REFERENCE   2  (bases 1 to 7925)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, Nat. Ins.
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..7925
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="15"
                     /tissue_type="stomach"
     gene            <1..>7925
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT     1454 a   2543 c   2691 g   1237 t
ORIGIN
```

Query Match 47.6%; Score 2055.4; DB 3; Length 7925;
Best Local Similarity 73.4%; Pred No. 0;
Matches 2833; Conservative 0; Mismatches 871; Indels 155; Gaps 10;









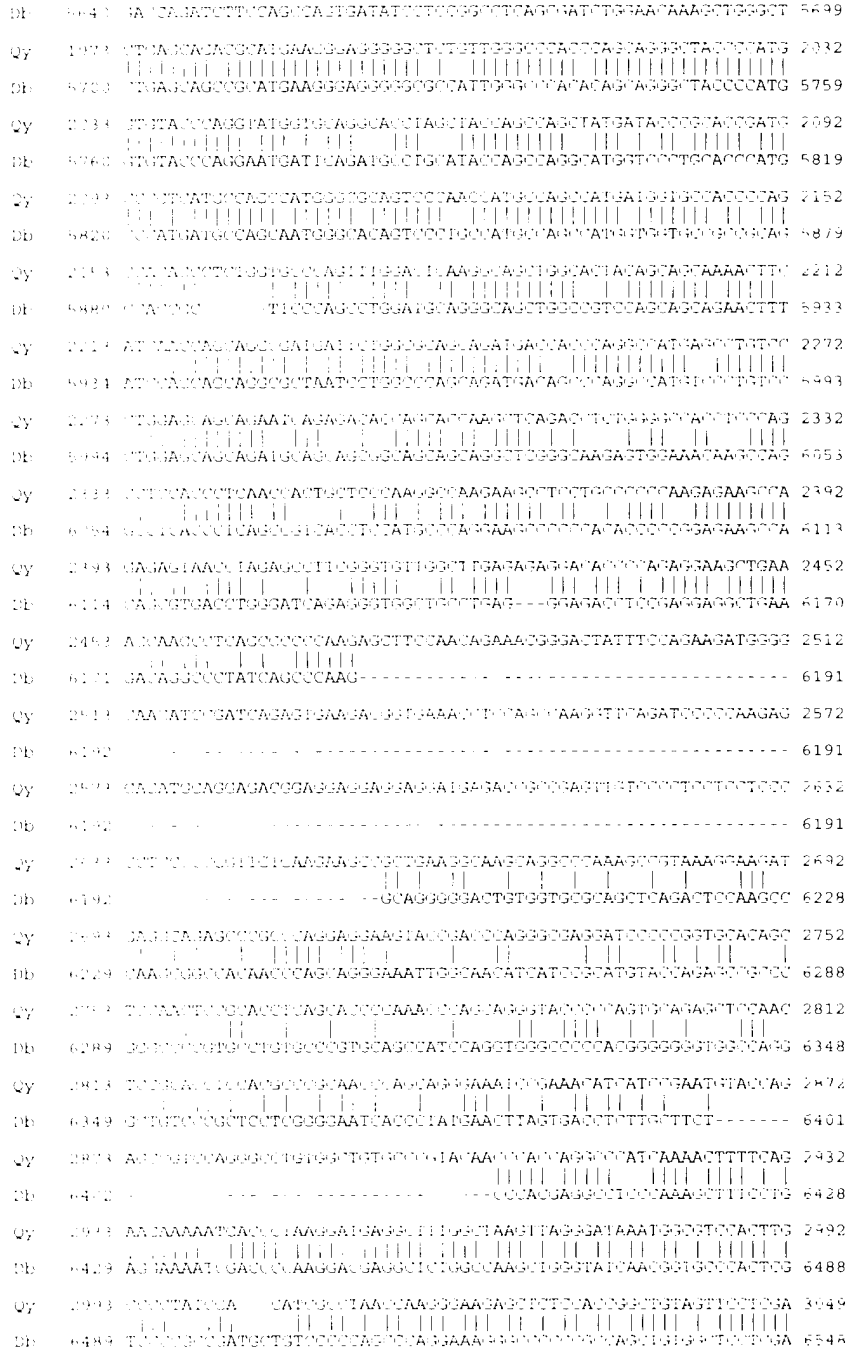
```
RESULT 5
AF418286
LOCUS       AF418286                7515 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens clone 1 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418286
VERSION     AF418286.1  GI:16303809
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7515)
  AUTHORS   Boger,F.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-488 (2001)
  MEDLINE   21960014
   PUBMED   11964073
REFERENCE   2  (bases 1 to 7515)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..7515
                     /organism="Homo sapiens"
```

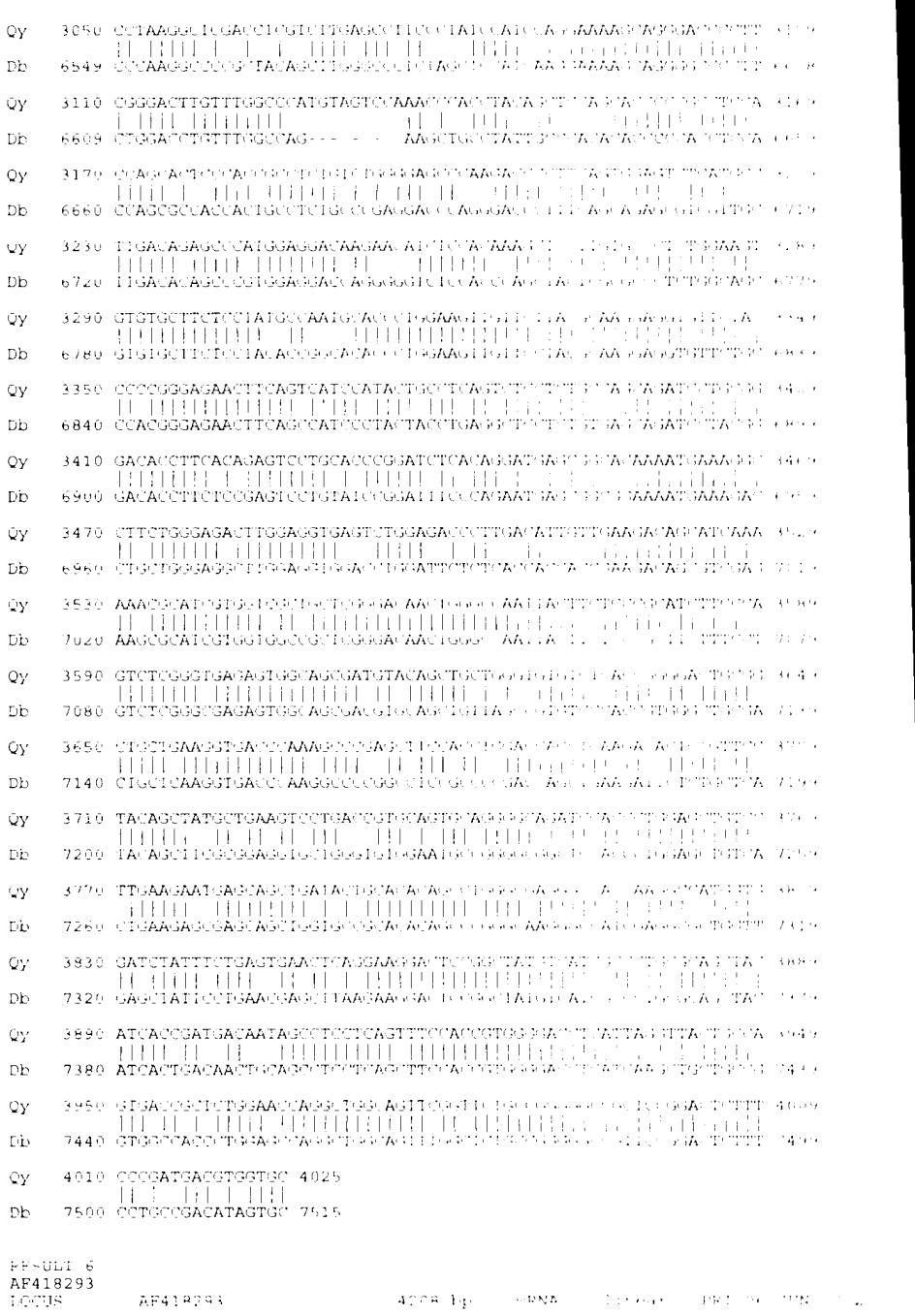
```
                     /db_xref "taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone "1"
                     /tissue_type="kidney"
     gene            <1..>7515
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT     1350 a    2469 c    2540 g    1156 t
ORIGIN

  Query Match             44.7%;  Score 1954.4;  DB 9;  Length 7515;
  Best Local Similarity   71.1%;  Pred. No. 0;
  Matches 2868;  Conservative    0;  Mismatches  836;  Indels  241;  Gaps   10;

Qy       2 GCTGGGACTGTCACCTACCAGGTTCAAAGTTTCATAAACAGAAACAGGGGCCAGCTGGAC 61
Db    3750 GCAGGGACTGTCACCTACCAGGTTCAAAGTTTTTAAACAGAAACAGGGATGAGCTGGAC 3809
Qy      62 CCCGCCGTTCTTGAAATCCTCAGGCAGAGCCAGTGCAGGTGACCTAGCCTTCCTTCAG 121
Db    3810 CCTGCTGTGTTGAAATCCTTGGCCAGAGCCAGC----------------TCCAG 3848
Qy     122 CTTATGGGCAGCCTTTCAAGAAGCAGAGCCCAGGCTGGGACTGAGCAAAACAAACCC 181
Db    3849 CTTTTGGGAAGCCTTTTAAGAAGCAGAGCCCAGGCAGGGAGGGCGAGGCAGAGCC 3908
Qy     182 ACATTGGCTTCTTCAATTCCAGCAGACCTGGGTGACTTGCTAGCTCGGCTAGCAGTAGG 241
Db    3909 ACATTGGCTTCTTCAATTCCAGCAGACCTTGGAAGACTTCATGGCCGGGCTGG--GCAGG 3965
Qy     242 GGCCATGTCTACGTCATCCACGTGTCTCAATCCTAGCCCTGGAAAGATCCCAGGCCTCTTG 301
Db    3966 AGCCAGTTCTATTTCATCCAGTGCCTCAGCCTAACCCTGGAAAGCTTCCAGGCCTCTTT 4025
Qy     302 GACGTGGGGCATGTGGCAGAGCAGTTGCGTCAGGCTGGCATCCTGGAGATCATAGGCAGC 361
Db    4026 GACGTGGGACATGTCACAGAGCAACTGCACCAGGCAGCTATTGGAGGCCGTGGCACC 4085
Qy     362 CGGAAGAGCACTTCCCCGTGCGAGTGTCCTTCAAGTCTTTCTGGCAAGGTTCCATGCC 421
Db    4086 CGCAGTGCCAACTTCCCCGTGCGTGTGCCCTTTGAGGCCTTCCTGGCCAGCTTCCAGGCC 4145
Qy     422 CTGGGGTCAGGGAGACAGAAAGCTGCTTTGACTAGGAGAGGTGTGGTGCCATCCTCAGT 481
Db    4146 CTGGGGTCAGAAGGGCAGGAAGACCTCTCTGACCGGAAGAAGTGTGGTGCCATCCTCAGT 4205
Qy     482 GAAGTGCTGGGGCAGAGTCAGCTGTATCATCTTGGAGTCAGCCAGGTCCTGCTGCAG 541
Db    4206 GAGGTGCTGGGGCGGAATCAGCTGTATCATCTTGGAGCCACAAGATCCTGCTGCAG 4265
Qy     542 GAACAGGGCTGGCAGCAGCTACAACAGCTGTGGCCTCAGCGGCGCTCACAGGCCTCCTC 601
Db    4266 GAGCAGGGCTGGCAGCAGCTGTGAGCAGTCCCAGGACAGCAGCGCTCCCAGGCCCTGCTC 4325
Qy     602 ACTCTGCAGCGTGGCCTCCCAGCCTGTACCA------CAGCGGCTCCGTCTCCTGCCGG 661
Db    4326 GAGCTGCAGCGCAGTCTCCAGAGTGTATCTCCCCAGCGG---- GTCCTGCCCGG 4379
Qy     662 ATGCAGGCTCGTGTGCCTCCTAGGCCAGGAAGCGATACCTCCAGCGGAGGTCAGCT 721
Db    4380 ATGCAGCTTCAGATGCGTGGCTTCCAGGCCAGAAGCGGTACCTCCGGCGGCGGCAGCT 4439
Qy     722 CTGGGACAGCTGAACACCATTCTCCTAGTGGCCCCGCCCCTGCTCCCGAGACGACAGAAG 781
Db    4440 TTGGGACAGCTGAACACCATCCTGCTGGTGGCCCAACCCCTGCTCCAGAGCGGGCAGAGA 4499
Qy     782 CTAGCTTCTCCAGCTGGAGCAG------CAGAGCGGGAGCCCTGGGGAAAGTCCCA 832
Db    4500 CTGCAGCTTGGGTCTCTCAGTGCTTGGCATAGCAGCGAAAGGGCTTGGAGAGAGTGGCA 4559
Qy     833 AATATGGACCTGGGTCGCTTAGAGATCCCCGCCCAGCTGGCTACTCTGCTGGAGAGGGCG 892

Db    4560 AGCATGGAGCTGGGGCGCTTGGAGATCCCGGCTGAGCTGGCCGTCATGCTGAAGACGGCG 4619
Qy     893 GAAGGCCACCAGGCCTTGCTGACGGGGAGCATCACAGATGCCTTCAACTAGCTCCC 952
Db    4620 GAAAGCCATCGGGAGGCCCTGGCTGGGAGCATCACAGGAGTCTTCAGGTTCCT 4679
Qy     953 GCCCGGCCCAGCCTGACTCTCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTA 1012
Db    4680 GCCCGGCCCAGCCTGACTCTCCCAGCAGACATTGACCTGTTCCCTTTCTCCAGCTTCGTC 4739
Qy    1013 TCCACCAGCTTTCAGAAGCCATTTCTGCCTCGACCAGGGCAGCCACTGGAGGAGCCCCTG 1072
Db    4740 GCCATCGGCTTTCAGGAACCGTCCCTGCCCAGGCCAGGGCAGCCACTGGCGAAGCCCCTA 4799
Qy    1073 AGGCGGTTACATGGCGAGAACCCTCACCAAGTTTTGGAGATCAACAGGGTGATGCTGCGG 1132
Db    4800 AGCCAGCTGGATGGAGACAACCCTCAGCTGCCTGGACATCAACAAAGTGATGCTGCGG 4859
Qy    1133 CTCCTGGGGGAAGGATCTCTGCAGCCTGGCAACAGCAGACCATGGCACGTTCCTCGTG 1192
Db    4860 CTCCTGGGGGAAGGATCCCTGGAGCTCCTGCAGAGCCAGATCATGGCACATACCCGTG 4919
Qy    1193 CAGCAGGCCAGCGACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAG 1252
Db    4920 CGGCAGGGGCAGTGGCCGGCCAGGGCTACGGAATGAGCTCTTTAGCCAGCTGGTGGCCCAG 4979
Qy    1253 CTGTGGGCAACTCAGATGAGCAACAGAATCAGCGTGGCTGGGCCCTAATGGTGATCCTG 1312
Db    4980 CTATGGCAGAACCCAGATGAACAGCAGAGCTAGCGTGGCTGGGCCCTCATGGCTGTTTTG 5039
Qy    1313 CTCAGCTCCTTTGCTCCCACAGCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATCTGAC 1372
Db    5040 CTCAGGGCCTTTCCCCCACTGCCCGTCCTACAGAAGCCACTGCTCAAGTTCGTGTCTGAC 5099
Qy    1373 CAGGCTCCCAGTGGCATGGCAGCCCTGTGCCAGCACAAGCTGTTAGCTGCCCTGGAGCAG 1432
Db    5100 CAGGGCCCCAGGGGCATGGCAGGGCTGTGCCAGCACAAGCTGCTGGGGGCCTTGGAGCAG 5159
Qy    1433 AGAGGGCTGGCTGCCATGGCTCGAGGTGCCAGCCACCCACACAACTTCAGTGGAAGGCT 1492
Db    5160 TCGCAGCTGGGCTCGGGGGGCCACTCGGGCAGTCCCCAGCCACCAGCTGGAGTGGCTGCCC 5219
Qy    1493 GGTTTACGTCGGGGCCGCATGGCGCTGGATGTGTTCACATTCAACGAGGAAAGCTACTCC 1552
Db    5220 GGATGGCGGCGGGGCCGCATGGCGCTGGATGTGTTCACCTTCAGCGGGAGCGCTACTCG 5279
Qy    1553 GGGGAAGTGGAATCCTGGACCAGGGGAAAGCAGTTTGGCAGGTGGATCCTACAGAGCAGA 1612
Db    5280 GGCGAAGTGGAGTCCTGGACCAGCGGAAGCAGCTGGCTGGGTGGATCCTGCAGAGCAGA 5339
Qy    1613 GGCCTGAAGGGCCGCTGTGGCTGCTTTGACTGCATTCTGGGGATGCTTGGCGT 1672
Db    5340 GGCCTGGAGGGCCTCCCGGGGCTGCTGTGTCACTGCACTCCAGGGACGCATGGCAG 5399
Qy    1673 GACTTGCCTGGCTGTGACTTTGTGTTGGACCTAATAGGTAGACTGAGGACTTGGGAGAC 1732
Db    5400 GACTTGCCTGGCTGGGACTTTGTGCTGGACCTGATAGGCAGACTGAGGACCTGGGGGAC 5459
Qy    1733 CCAGCTGGTCCCCACAACTAGCAGCATCACTGCTGTTGGTTAGCTGAGACCATCCCTCCA 1792
Db    5460 CCAGCTGGCCCCGCAGCTACCGCATCACTGCCGTTGGCTCAGCGGAGGCCATTCCCCTT 5519
Qy    1793 GCCCCTGGTGTCCAGGCTGCTTCCTGCTGCAGGACTCCCTCCAGGTCCAGCCCCAATA 1852
Db    5520 GCCCCTGGCATTCAGGCCCCCTCACTGCCCCCAGGACCCCCTCCAGGTCCAGCCCCAACG 5579
Qy    1853 CTGGCCAGCAGCCGCCGTCCGGCGAGTCAGTAAGCCTCAGAACCTGGATGGTTTCGTG 1912
Db    5580 CTGCCCAGCAGGGACCACAGGGAGTCCACAGGTCAGGAGCCTCGACGGCTTCCTG 5639
Qy    1913 GACCAGCTTCTTGAAGCAGCGTCGCTCCGTTCAGTGACTTGGAACAAGCCTGGGCC 1972
```





LOCUS AF418293 4208 bp mRNA

```
EFINITION  Homo sapiens clone cDNA2 myosin XVPP (MYO15B) pseudogene, partial
           sequence.
CCESSION   AF418293
ERSION     AF418293.1  GI:16903816
EYWORDS    .
OURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE   1  (bases 1 to 4208)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVPP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21963214
   PUBMED   11964073
EFERENCE   2  (bases 1 to 4208)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
EATURES                  Location/Qualifiers
     source              1..4208
                         /organism="Homo sapiens"
                         /db_xref="taxon:9606"
                         /chromosome="17"
                         /map="17q24-q25"
                         /clone="cDNA2"
                         /tissue_type="kidney; stomach"
     gene                <1..>4208
                         /gene="MYO15B"
                         /note="myosin XVPP"
                         /pseudo
ASE COUNT      862 a   1386 c   1275 g    685 t
RIGIN

  Query Match           34.5%;  Score 1511.4;  DB 9,  Length 4208,
  Best Local Similarity  73.6%;  Pred. No. 0;
  Matches 2056;  Conservative    0;  Mismatches  656;  Indels   81,  Gaps    7;

Qy   1344 AGAAGCCACTGCTCAAATTTGTATCTGACCAGGCTCCCAGTGGCATGGCAGCCCTGTGCC 1403
Db      2 AGAAGCCACTGCTCAAGTTCGTGTCTGACCAGGCCCCCAGGGCCATGGCAGCCCTGTGCC 61

Qy   1404 AGCACAAGCTGTTAGGTGCCCTGGAGCAGACACCGCTGGCTCCCATGGCTTCGAGGTCGC 1463
Db     62 AGCACAAGCTGCTGGGGGCTTTGGAGTAGTCCCAGCTGGGCTCGGGGCCACTCCGGCCC 121

Qy   1464 ACCCACCCACACAACTTGAGTGGAAGGCTGGTTTACTGCGGGCCGCATGGCGCTGGATG 1523
Db    122 ACCCCCCACACCAACTTGGAGTGGCAGGCCGGGGGCCGCATGGCGCTGGATG 181

Qy   1524 TGTTCACATTCAACGAGGAAAGCTACTCCGCGGAAGTGGAATCCTGGACCACCGGAGAGC 1583
Db    182 TGTTCACCTTCAGCGAGGAGTACTCACTGCGCCAGTTGGAGTCTTGGACCACCGGGGAGC 241

Qy   1584 AGTTTGCAGGGTGGATCCTACAGAGC---------------------------------- 1609
Db    242 AGCTGGTTAGGTGATCCTGCAGAGCAGTTCAGAAAAGAAATGCCCATGCTGGAGCCTTT 301

Qy   1610 ------------------AGAGGCCTGGAGGCGCCCCCTCGTGGCTGGTCTGTGTCACT 1650
                           |||||||||||||||||| || ||||||||||||||||||||
Db    302 GTGGAAGTGGAGGGCCGTGAGAGGCCTGGAGCGCCCCTCCCCGGGGCTGGTCTGTGTCACT 361

Qy   1651 GCATTCTGGGGATGCTTGGGTGACTTGCCTGGCTGTGACTTTGTGTTGGACCTAATAGG 1710
Db    362 GCACTCCAGGGACGCATGGCAGGACTTGGTTGGCTGCGACTTTGTGCTGGAGCTGATCAG 421

Qy   1711 CCAGACTGAGGACTTTGGAGACCCAGCTGGTCCCCACAACTACCCCATCACTCCTCTTGT 1770
Db    422 CCAGACTGAGGACCTGGGGGACCCAGCTGGCCCCCCAGCTACCCCATCACTCCTCTTGT 481

Qy   1771 TTTAGCTGAGAGCATCCCTCCAGAGCCTGCTGTCCAGGCTCCTTCCCTGCCCCCAGGACT 1830

Db    482 CTCAGCCGAGGCCATTCCCCTTGCCCTGCTATTCAGCCCCCTCACTGCCCCCAGGACC 541

Qy   1831 CCCTCCAGGTCCAGCCCCAATACTGGCCAGCAGCCGCCCTCCGGGCGAGGCCAGTAAGCC 1890
Db    542 CCCTCCAGGTCCAGCCCCAAGGCTGCCTAGCAGGAGCCACACAGCCAGGCCCAGAGGTC 601

Qy   1891 TGAGAACCTGGATGGTTTCGTGGACCACCTCTTTGAACCAGCGGCTCGCTCCGGGTTTCAG 1950
Db    602 AGGGAGCCTGGACGGCTTCCTGGACAGATTTCAGCCAGTGATATCCTCCGCCTCAG 661

Qy   1951 TGATCTGGAACAAGGCTGGGCCCTGACAGAAGCATGAAGGAGAGGGGCTCTGTTGGGCC 2010
Db    662 CGATCTGGAACAAAGCTGGGCTCTGAGCAGCCGCATGAAGGAGGGGCCCCATTGGGCC 721

Qy   2011 CACCCAGCAGGGCTACCCCATGGTGTAGCCAGGTATGGTCCAGCCACCTAGCTACCAGCC 2070
Db    722 CACACAGCAGGGCTACCCCATGGTGTAGCCAGGAATGATTCAGATGCCTGCATACCAGCC 781

Qy   2071 AGCTATGATACCCGCACCGATGCCCGTCATGCCAGCCATTGGGCCAGTCCCAACCATGCC 2130
Db    782 AGGCATGGTCCCTGCACCCATGCCCATGATGCCAGCAATGGGCACAGTCCCTGCCATGCC 841

Qy   2131 AGCCATGATGGTGCCACCCCAGCCACAGCCTCTCGTGCCCAGTTTGAGTTAAGGCAGCT 2190
Db    842 AGCCATGGTGGTGCCGCCGCCAGCCACCGC - - -TTCCCAGCCTGGATCCAGGGCAGCT 895

Qy   2191 GGCACTACAGCAGCAAAACTTCATCAAGCAGCAGGGGATGATTGTGGCGCAGCAGATGAC 2250
Db    896 GGCCGTCCAGCAGCAGAACTTATCCAGCAGCAGCGCTAATCCTGGCCAGCAGATGAC 955

Qy   2251 CAGCCAGGCCATGAGCCTGTCCCTGGAGCAGCAGAATCAGACAGCAGCAGCACCAAGCTCA 2310
Db    956 AGCCCAGGCCATGTCCCTGTCCCTGGAGCAGCAGATGTAGCAGCGGCAGCAGGCTCG 1015

Qy   2311 GACCTCTGGGGCCACCTCCCAGCCTCCAGCCTGAACCACTGCTCCCAAGGCCAAGAAGCC 2370
Db   1016 GGCCTCCGAGGCTGCGTCCCAGCCTCAGCCTCAGCCGTCACCTCCAAGCCCAGGAAGCC 1075

Qy   2371 TCCTGCCCCCCAAGAGAAGCCAGAGACTAACCTAGACCTTCCCCCCTTCGCCTTGAGAGA 2430
Db   1076 CCCCACACCCCCGGACAAGCCACAGCCTGACCTGGGATCAGACGGTGGCTGCCTGAG--- 1132

Qy   2431 GGAGAGCCCAGAGGAAGCCGAAAAGAAGCCTCAGCCCCCAAGAGCTTCCAACAGAAACG 2490
Db   1133 GGAGACCTCCGAGGAGGCTGAAGACAGGCCTATCAGCCCAAGACCTTCCAGCAGAAACG 1192

Qy   2491 GGACTATTTCCAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACGGTGAAACCTCCAGC 2550
Db   1193 GAACTATTTCCAGAAGATGGGGCAGCCACAGATCACAGTGAGGACGATGAAGCCCCCGGC 1252

Qy   2551 AAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGATGAGACCGC 2610
Db   1253 CAAGGTCCACATCCCCCAGGGGGAAGCCACGAGGAGGAGGAGGACGACGAGGAGGAGGA 1312

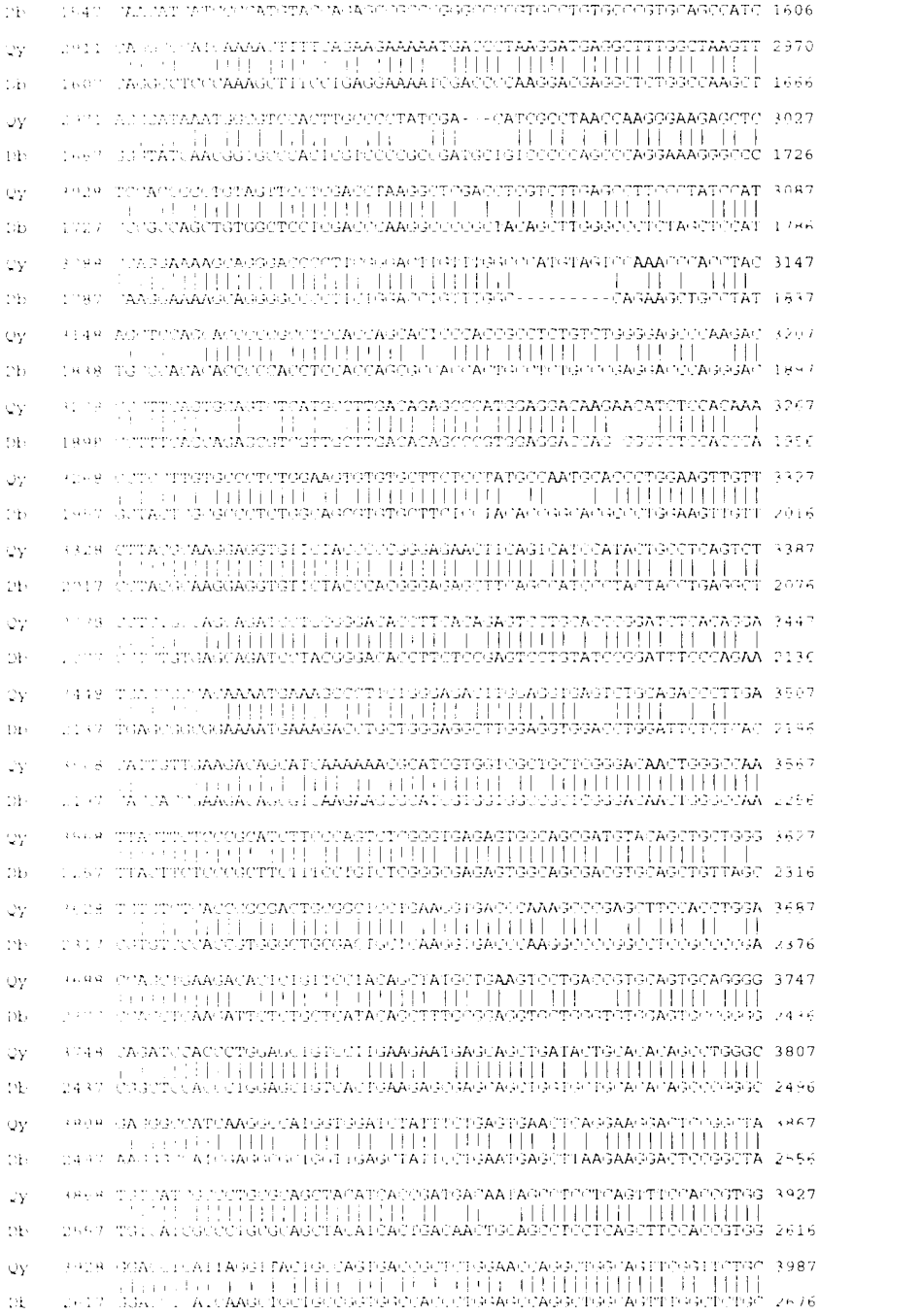
Qy   2611 GGAGTTGTCCCCTCCTCCTCCCCCTCCCCCGTTGTGAAGAAGCCCTGAAGGCAAGCAG 2670
Db   1313 GGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGTGCCGTCCCCTC - ---CTCCTCC 1366

Qy   2671 GCCCAAAGCCGTAAAGGAAGATGAGGCAGAGCCCGCCCAGGAGGAAGTACCGACCCAGGG 2730
Db   1367 CCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGTTCCAAAAGAGGCTGAGGC 1426

Qy   2731 CGAGGATCCCCCGGTGCACAGCTCCAACTCCCCACCTCAGCACCCCAAACCCAGCAGGGT 2790
Db   1427 TGAGCCAGCCAAGGAGACAGCAGCCAAGGCCATGGCAAGGCCAGCCCAAGGCAGGGG 1486

Qy   2791 ACCCCAGTGCAGAGCTCCAACTCCCACCTCCACCCCCAACCCAGCAGGGAAATCCG 2850
Db   1487 GACTGTGGTGGGAGCTCAGACCCAAGCCCAAGCCCCACAACCCAGCAGGGAAATTGG 1546

Qy   2851 AAACATCATCCGAATGTACTAGAGCCTCCAGGGCCTGTGGCTGTGCCCGTACAACCCAC 2910
```



Qy 3988 CGGGGGCCGCTCCGGACTCTTTCCCGATGACTGGTCAGCTACCTCCTGACTT 4047
Db 2677 CGGGGGCCGCTCCGGACCCTTTCCTGCCGACATCAGCGACCTCCTGACTT 2726

Qy 4048 CTCCTTTTCCCTGGGAAAGAGAAACAGCTGGCA 4080
Db 2737 TTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 2769

RESULT 7
AF418292
LOCUS AF418292 4602 bp mRNA linear PRI 09-JUN-2001
DEFINITION Homo sapiens clone cDNA1 myosin XVB (MY15B) pseudogene, partial sequence.
ACCESSION AF418292
VERSION AF418292.1 GI:16304815
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4602)
AUTHORS Boger,E.T., Sellers,J.R. and Friedman,T.B.
TITLE Human myosin XVBP is a transcribed pseudogene
JOURNAL J. Muscle Res. Cell. Motil. 22 (5), 477-484 (2001)
MEDLINE 21960014
PUBMED 11964073
REFERENCE 2 (bases 1 to 4602)
AUTHORS Boger,E.T., Sellers,J.R. and Friedman,T.B.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, 5 Research Court, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1..4602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q24-q25"
/clone="cDNA1"
/tissue_type="kidney; stomach"
gene <1..>4602
/gene="MYO15B"
/note "myosin XVBP"
/pseudo
BASE COUNT 927 a 1495 c 1405 g 775 t
ORIGIN

Query Match 34.5%; Score 1511.4; DB 9; Length 4602;
Best Local Similarity 73.6%; Pred. No. 0;
Matches 2056; Conservative 0; Mismatches 659; Indels 81; Gaps 7;

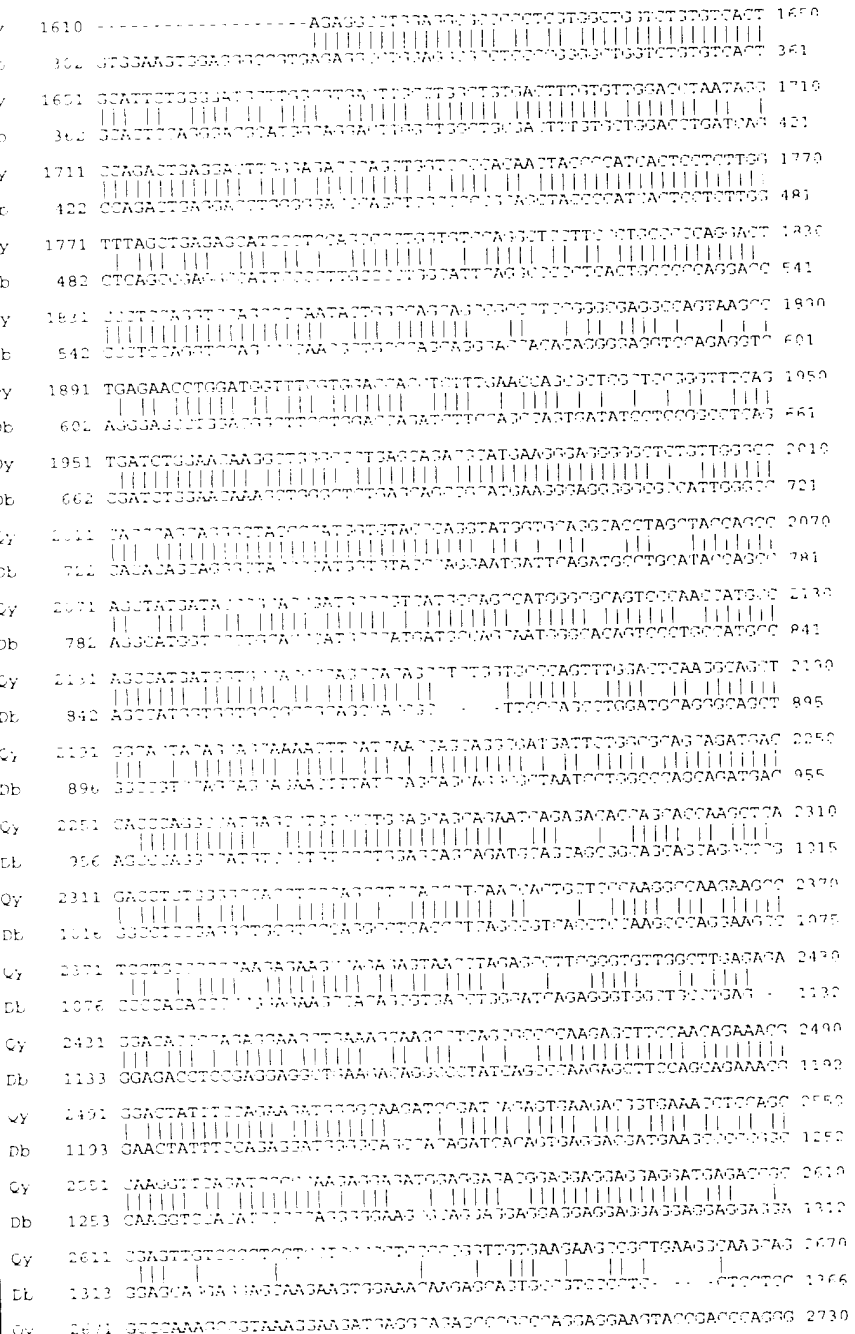
Qy 1344 AGAAGCCACTGCTCAAATTTGTATCTTACTCAGGTCTTCAGTCCATCTTAGCCTTTCT 1403
Db 2 AGAAGCCACTGCTCAAGTTCGTGTCTGACCAGGCTTCAGGCCATCTTAGCCTTTCT 61

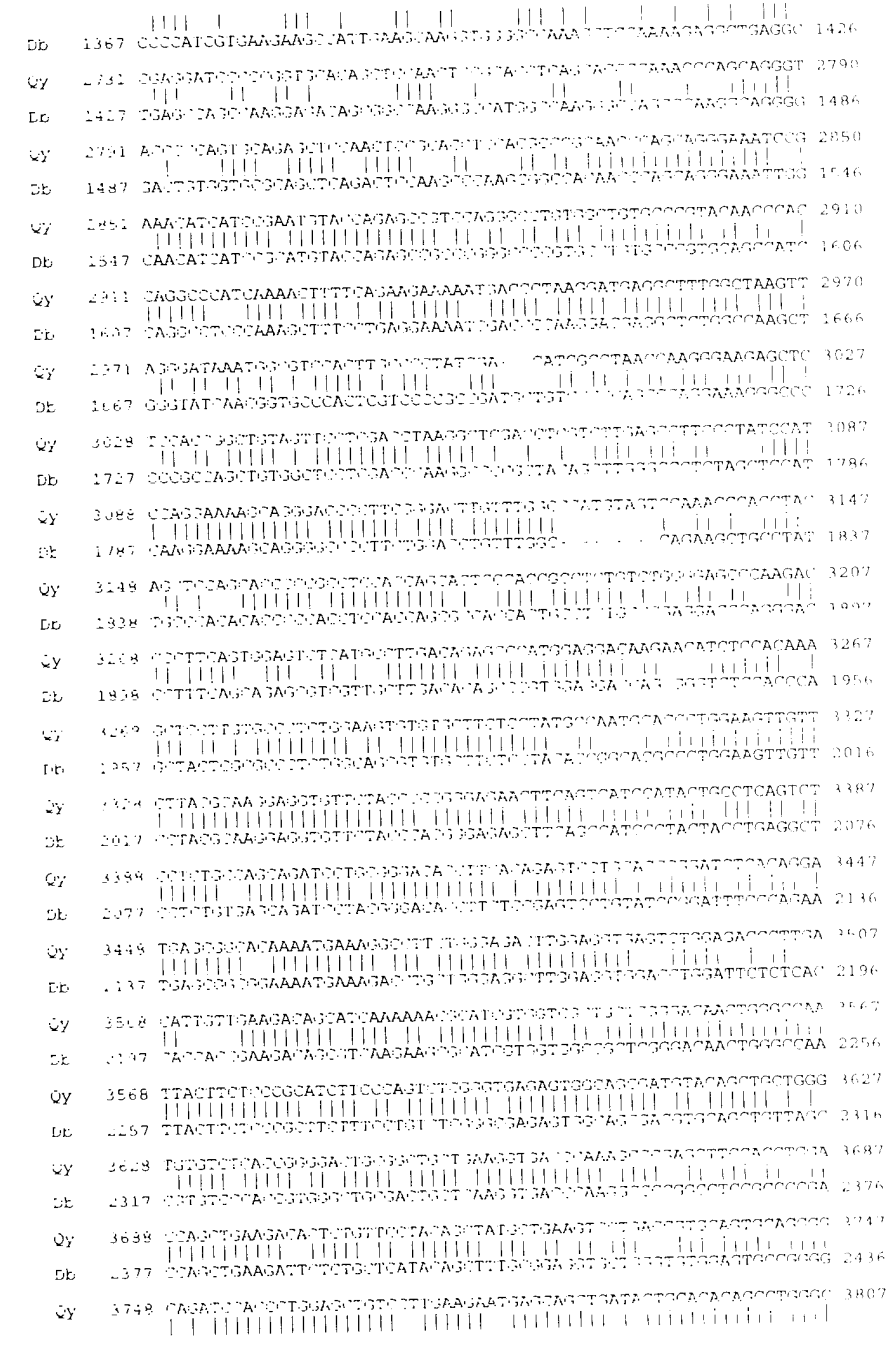
Qy 1404 AGCACAAGCTGTTAGGTGCCCTGGAGCAGACACCGCTGGCTCCCATGGTTGAGTCCC 1463
Db 62 AGCACAAGCTGCTGGGGGGCCTGGAGCAGTCGCAGCTGGCTT-----ACTCCCCC 121

Qy 1464 ACCCACCCACACAACTTGAGTGGAAGGCTGGTTTACCTCCTGGGTCTCATCGCTGATCTC 1523
Db 122 ACCCCCCGACCCAGCTGGAGTGGCTGGCCGGATGGCGGCGGAGTCGATGGTCTGGATCT 181

Qy 1524 TGTTCACATTCAACGAGGAAAGCTACTCCTGCAAGCCGAA-CTGAACATCGCAGAGC 1580
Db 182 TGTTCACCTTCAGCGAGGAGTGCTACTCCGCGAGTCGTGTTTGAACTCTCGCAGC 241

Qy 1584 AGTTGCAGGGTGGATCCTACAGAGC---- 1609
Db 242 AGCTGCTGGGTGGATCCTGCAGAGCAGTCAGAAAGAAATGCCATCTCTGAGCCCTT 301





```
Db    2437 AGGCTCCACCCTGGAGCCTCTCACTGAAGAGCGAGCAGCTGGTGCTGCACACAGCCCGGGC 2496

Qy    4029 GAAGGCCATCAAGGCCATGGTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTA 3867

Db    2497 AAAGGCCATCAAGGCGCTGGTTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTA 2556

Qy    3866 TGTCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCAGTTTCGACGGTGG 3927

Db    2557 TGTCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCAGTTTCGACGGTGG 2616

Qy    3926 GAACTCATTAGCTTACTGCCTAGTGACGGCTCTGGAACCAGGGTGGCAGTTTGGTTCTGC 3987

Db    2617 GAACTCATCAAGGTGCTGCCGGTGGCCACCCTGGAGCTAGGCTGGCAGTTTGGTCTTGC 2676

Qy    3986 CGTGGTCTGCTCCGGACTCTTTCCCGATGATGATGAGCAGCAGTGGTCGCCCCGACCT 4047

Db    2677 AGGGGCCGTTCCGGACCCTTCCTGCCGACATAGTGCAGCTGCTCCCTGACTT 2736

Qy    4046 TTCTTTTCCCTGGGAAAGAGAAACAGCTGGCA 4080

Db    2737 TTCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 2769


RESULT 3
AF418289
LOCUS       AF418289                5709 bp    mRNA    linear   PRI 06-JUN-2002
DEFINITION  Homo sapiens clone 11 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418289
VERSION     AF418289.1  GI:21162812
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5709)
  AUTHORS   Roger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21962014
    PUBMED  11964073
REFERENCE   2  (bases 1 to 5709)
  AUTHORS   Roger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..5709
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="11"
                     /tissue_type="stomach"
     gene            <1..>5709
                     /gene "MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT     1641 a   1840 c   2017 g    821 t
ORIGIN

  Query Match             25.6%;  Score 1119.4;  DB 9;  Length 5709;
  Best Local Similarity   75.2%;  Pred. No. 3.6e-245;
  Matches 1480;  Conservative    0;  Mismatches  416;  Indels   71;  Gaps    4;

Qy     685 ATAGGCAAGAAGCGATATCTCCAGGGGAGGTCAGCTCTGGGACAGCTGAACACCATTCT 744

Db    3753 ATTGAAAGGAAGCGGTACCTCCGCCGTGGGCAGCTTTGGGACAGCTGAACACCATCCT 3811

Qy     745 CTAATGGGCCGGCCCTGTCTCCAGGACACGACACAAGCTACGGTGTCCCCCTGGCCCG-- 802

Db    3812 GCTGCTGCCCAGCCCCTGCCCCAGAGGCGGCAGAGACTGCAGCTTGGGGCTTGCCACAGG 3871

Qy     803 ------CACAGCGGGGAGCCCTGGGGCAAAGTGCTCAAACTATGAACTGGGTCCTTACA 856

Db    3872 CTGGCATAGCAGCGAAAGGGCCTTGGAGACAGTGCTAACCTACTGAATTGGGTCCTTACA 3931

Qy     856 GATCCCCGGCCCAGCTGGCTACTCTGCTGGAGAAGGCGAAGTCTACTATGTCTTCATGA 916

Db    3932 GATCCCGGCCTGAGCTGGCCGTCATGCTGAAGACGGCTGAAAGCTACTCTGACGTCGTCT 3991

Qy     916 GGGGAGCATCACAGAGTGCCTGCCACGTGAGGTCCCCGACTGGTCCAGCTTGACTTCTCC 976

Db    3992 TGGGAGCATCACCGAGTGCCTGCCGCCTGAGGTCCTGCCCGACTGGTCCAGCTTGACTTCTCC 4051

Qy     976 TCCAGACATTGACCAGTTCCCCTTCCCCAGTTTCGACTACTATTCTTCAGAACCATCT 1035

Db    4052 AGCAGACATTGACCTGTTCCCTTTCCCCAGTTTCGACTACTATTCTTCAGAACCATCT 4111

Qy    1036 TCTGCCTCGACCAGGGCAGCCACTGGACGAGCCCCTGACTGCTTTACATGGCGACAACCT 1095

Db    4112 CCTGCCCAGGCCAGGGCAGCCACTGGCGAAGCCCCTAACCTACTTGACATGGACAACCT 4171

Qy    1096 TCAGCAGGCTCTGGAGATCAACAAGTGATGCTTCCCCTCCTTTTACATGGACAACCT 1155

Db    4172 TCAGCGTGCCCTGGACATCAACAAAGTGATGCTTCCCCTCCTTTTACATGGACAACCT 4231

Qy    1156 GTCCTGGCAGAGGCAGATCATGGGCGTATACTTGGTGTCCCTAGTTTCTCCCTCAGT 1215

Db    4232 GTCCTGGCAGAGGCAGATCATGGGCGTATACTTGGTGTCCCTAGTTTCTCCCTCAGT 4291

Qy    1216 ACTCCGAGATGACGTTCTTCAGCCAGCTGGTGGCCAGTTATCTTGAAGTCCAGATGAAGA 1275

Db    4292 GCTGCGGAATGAGCTCTTTAGCCAGCTGGTGGCCAGTTATCTTGAAGTCCAGATGAAGA 4351

Qy    1276 ACGAAATCAGCCTCCTCCCCTCCAAGCGTGATCCTGCCAGCCTTCTCCACAC 1335

Db    4352 GCAGAGCCAGCGTGGCTGGGGCCCTCATGGTGTTTTG 4411

Qy    1336 TGCCTCAGAAGCTCACTCCTCAAATTCTCTATCTCAGTCAGTCTTCTCCAG 1395

Db    4412 TGTCCTACAGAAGCCACTGCTCAAGTTCCTGTCTGACCAGCTCCAGGGCAGTCTTCCAG 4471

Qy    1396 CCTGTGCCAGCACAAGCTGTTACCTGCCCTGGAGCAGACCCTCCTCCAGCTTC 1455

Db    4472 CCTGTGCCAGCACAAGCTGTTACCTGCCCTGGAGCAGACCCTCCTCCAGCTTC 4531

Qy    1456 GACTTCCACTCCACCCACTCAACTTCACTCCGAAGCCTCTCTTCACTTCCCCCACCGC 1515

Db    4532 TCTGGCCCACCCCCGGACCCACCCTCCCTGCCCTGACATCCCCGCCATGGC 4591

Qy    1516 GCTGGATGTGTTCACATTCAACCAGAAAGCTACTCCCCCGAAGTCCAACCCTGGACTCA 1575

Db    4592 GCTGGATGTGTTCACCTTCAGCGAGGAGTGCTACTCCCCGAAGTCCAACCCTGGACTCA 4651

Qy    1576 GCGAGAGCAGTTTGCAGGGTGGATCCTACAGAGC                    1610

Db    4652 CGGGGAGCAGCTGGCTGGGTGGATCCTCACAGACAAGCAGCAAAGAAGTCCATCTG 4711

Qy    1610 --  -  -  -  -  -  -  -  -  -  -  AGAGGCCTCTGAAAGCCCCCCCCCCTCCTCCT 1643

Db    4712 GAGCCCTTGTGGAAGTGGAGGCCGTGAGAGGCCTTGAAGTCTTTTTTTTTTTTT 4771

Qy    1644 GTGTCACTGCACTTCTGGGGATGTCTGGCCTGATCTTCCTCCCCCTTTGAC 1703

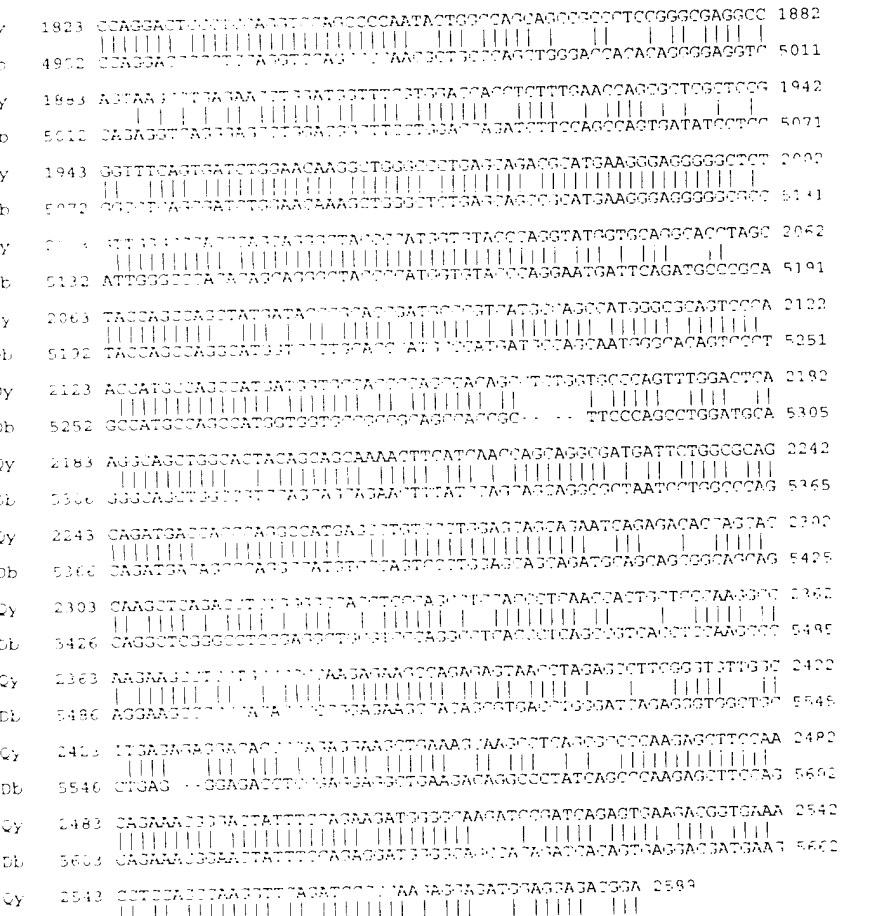
Db    4772 GTGTCACTGCACTTCCAGGCAGCATCTCAGCAGTCTTCCTCCCCCTTTGAC 4831

Qy    1704 CTAATAGGCCAGACTGAGGACTTGGGAGACCCAGCTGGTCCCCCACAATACCCCATCACT 1763

Db    4832 CTGATCAGCCAGACTGAGGACTGGGGGAGCCAGCTGGTCCCCCACAATACCCCATCACT 4891

Qy    1764 CCTCTTGGTTTAGCTGAGAGCATCCCTCAGCCCCTGGTCT-ACCCCTTCCCCTGCT 1821

Db    4892 CCTCTTGGCTCAGCCGAGGCCATTCCCTTTGCCCCTGGCATTCACCCCCTCACTGCCT 4951
```

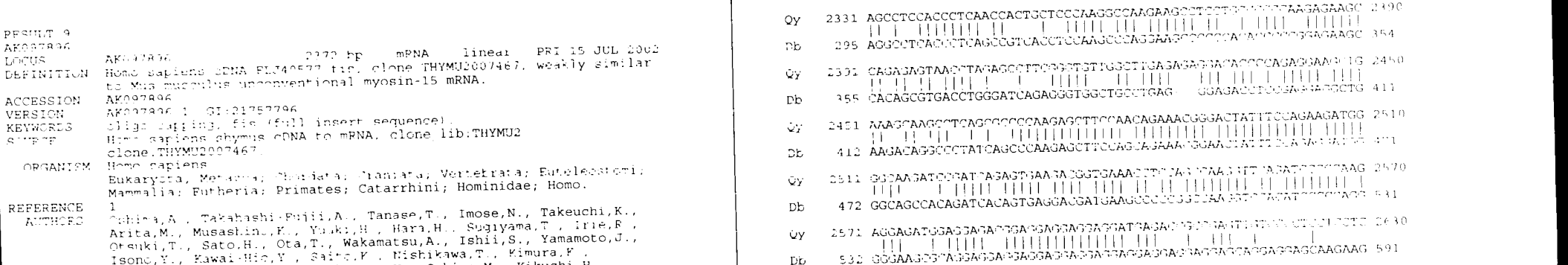


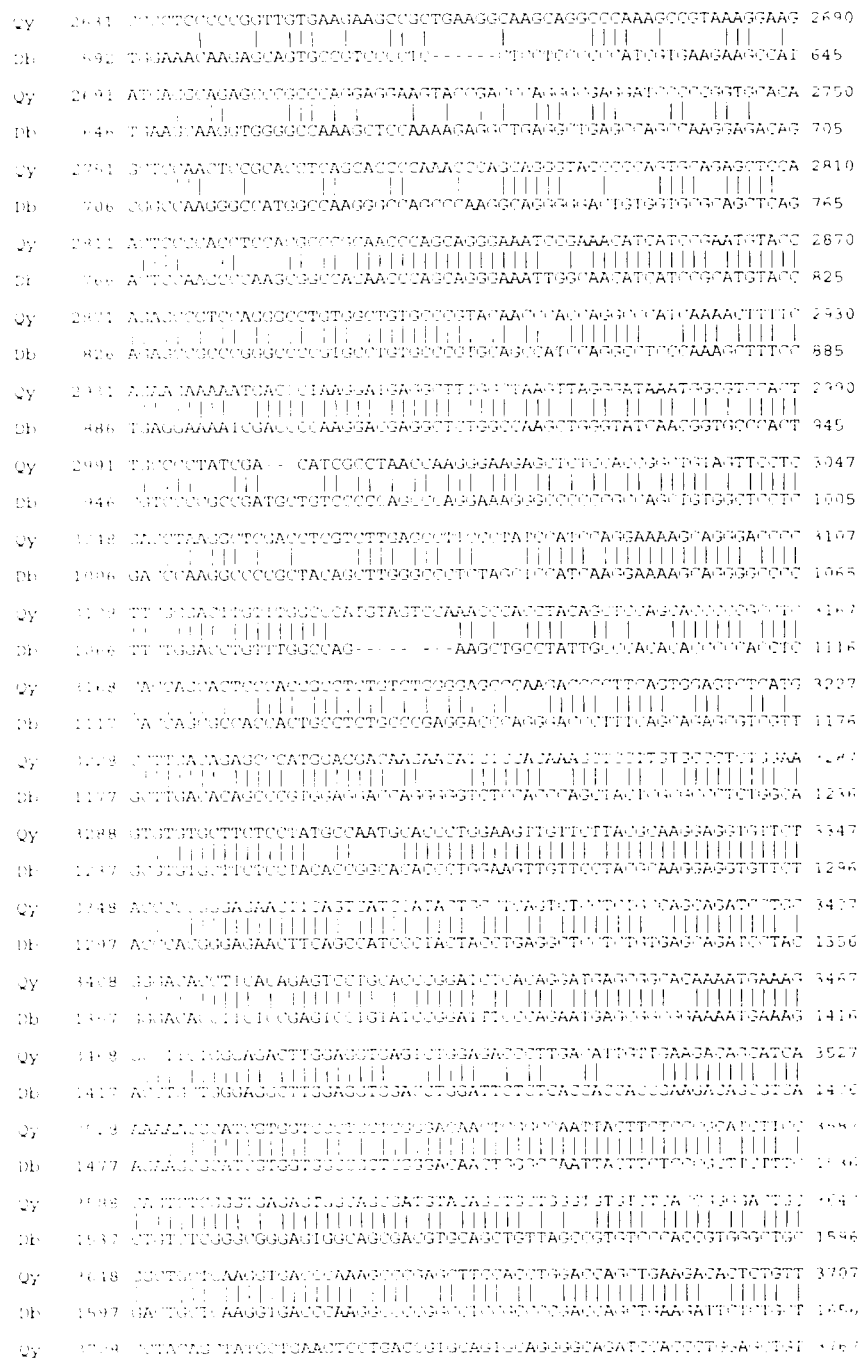
RESULT 9
AK007896
LOCUS AK007896 2372 bp mRNA linear PRI 15 JUL 2002
DEFINITION Homo sapiens cDNA FLJ40577 fis, clone THYMU2007467, weakly similar to Mus musculus unconventional myosin-15 mRNA.
ACCESSION AK007896
VERSION AK007896.1 GI:21757796
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens thymus cDNA to mRNA, clone lib:THYMU2 clone:THYMU2007467.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ishii,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2372)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3925, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan. cDNA full insert sequencing: Research Association for Biotechnology (RAB). cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

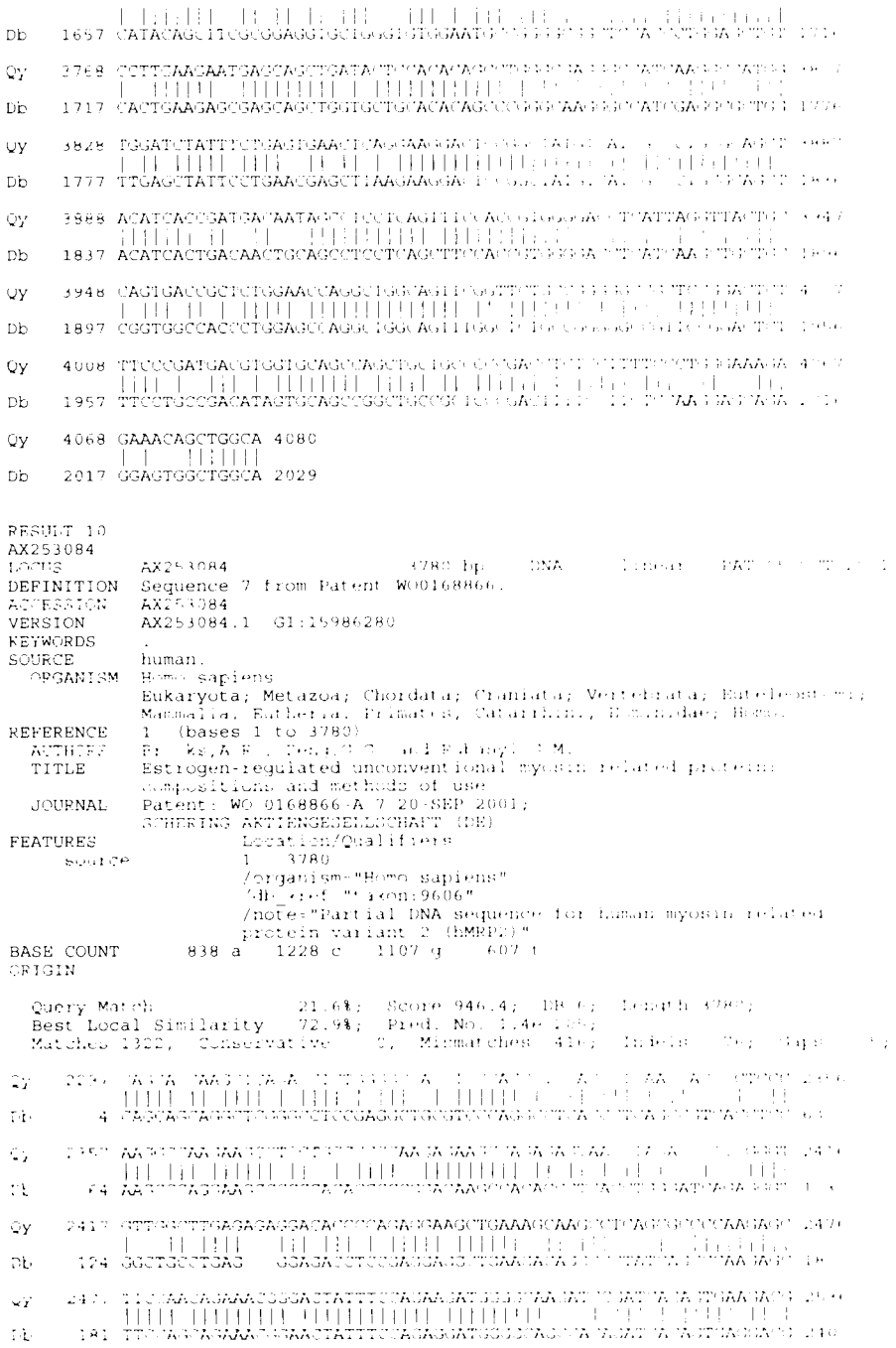
FEATURES Location/Qualifiers
source 1..2372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYMU2007467"
/tissue_type="thymus"
/clone_lib="THYMU2"
/note="cloning vector: pME18SFL3"

BASE COUNT 486 a 780 c 715 g 391 t
ORIGIN

Query Match 26.4%; Score 1110; DB 3; Length 2372;
Best Local Similarity 73.2%; Pred. No. 3.9e-243;
Matches 1662; Conservative 0; Mismatches 524; Indels 27; Gaps 5;

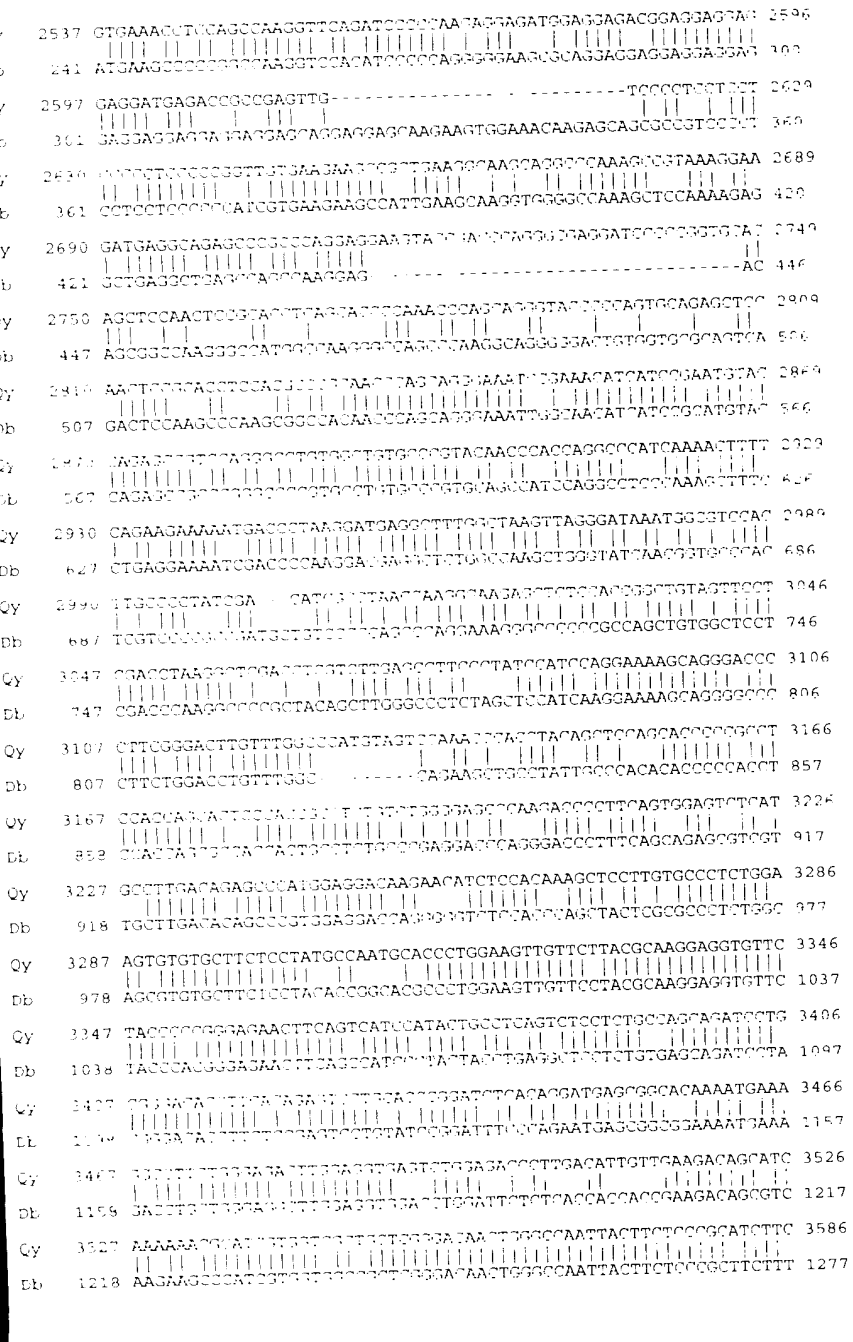


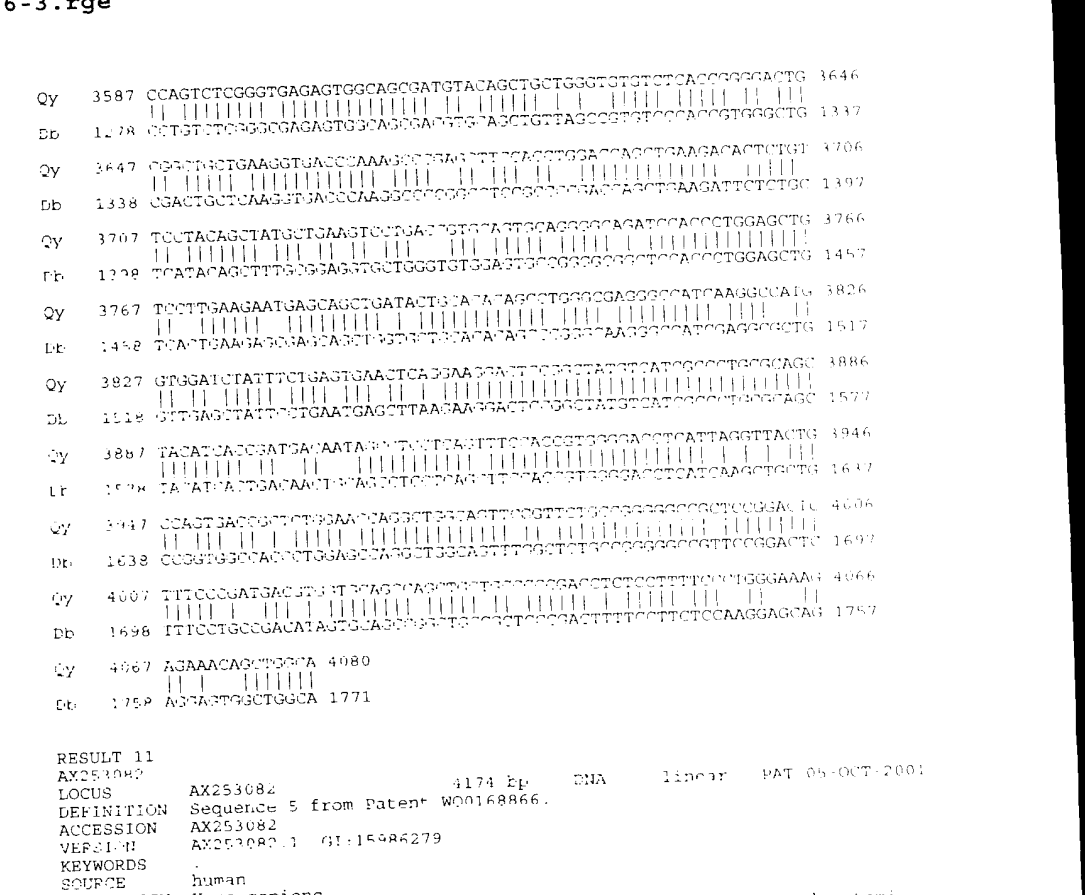




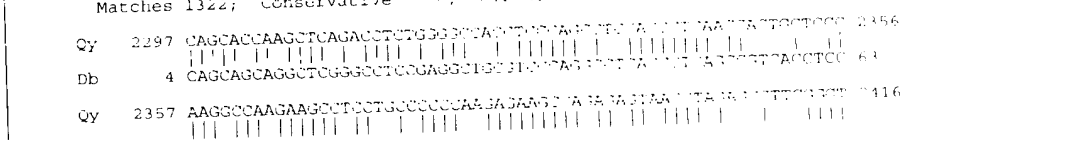
RESULT 10
AX253084
LOCUS AX253084 3780 bp DNA linear PAT
DEFINITION Sequence 7 from Patent WO0168866.
ACCESSION AX253084
VERSION AX253084.1 GI:15986280
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3780)
AUTHORS [illegible]
TITLE Estrogen-regulated unconventional myosin related proteins: compositions and methods of use
JOURNAL Patent: WO 0168866-A 7 20-SEP 2001;
SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1 3780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Partial DNA sequence for human myosin related protein variant 2 (hMRP2)"
BASE COUNT 838 a 1228 c 1107 g 607 t
ORIGIN

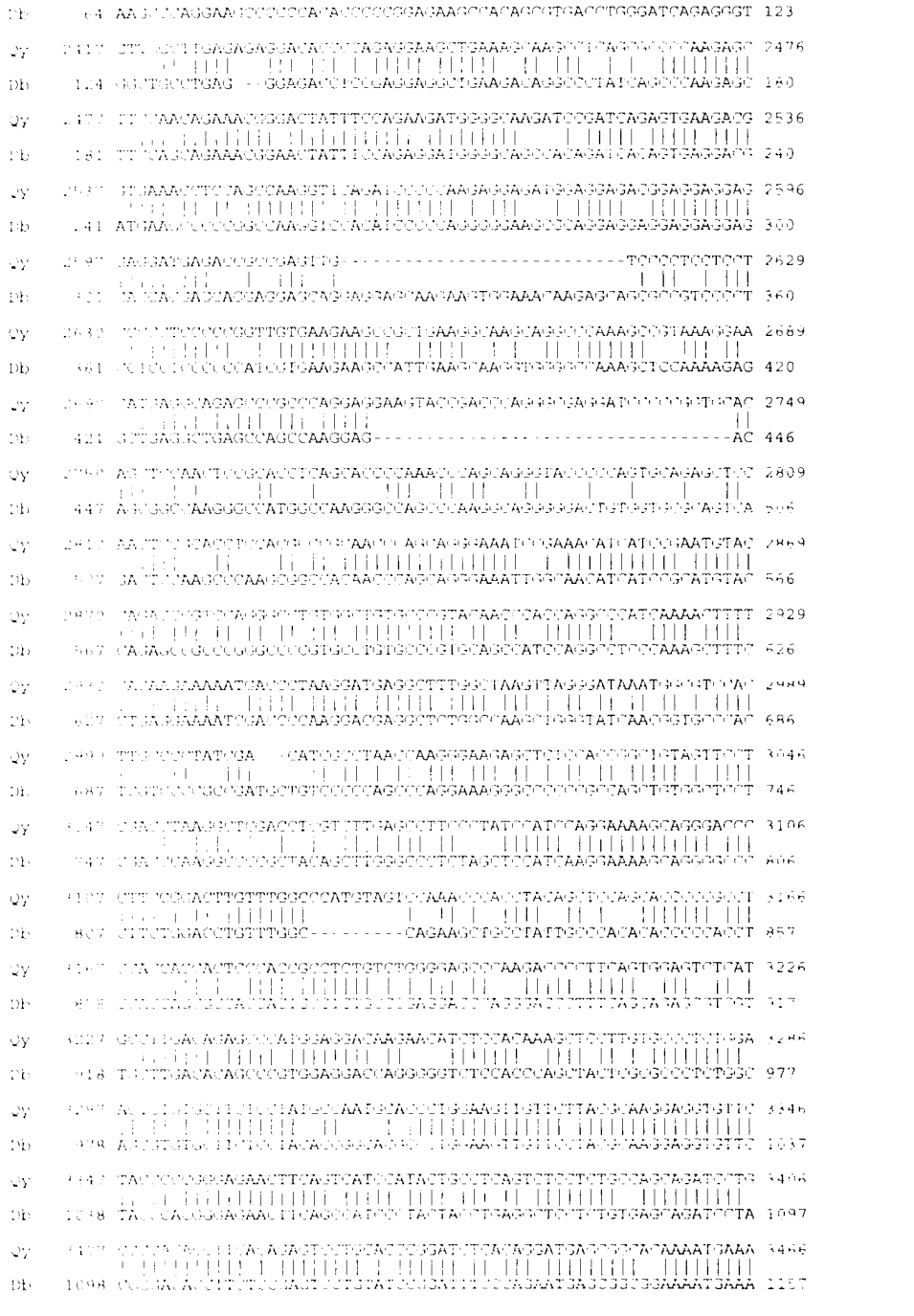
Query Match 21.6%; Score 946.4; DB 6; Length 3780;
Best Local Similarity 72.9%; Pred. No. 1.4e-155;
Matches 1322; Conservative 0; Mismatches 410; Indels [illegible]; Gaps [illegible];

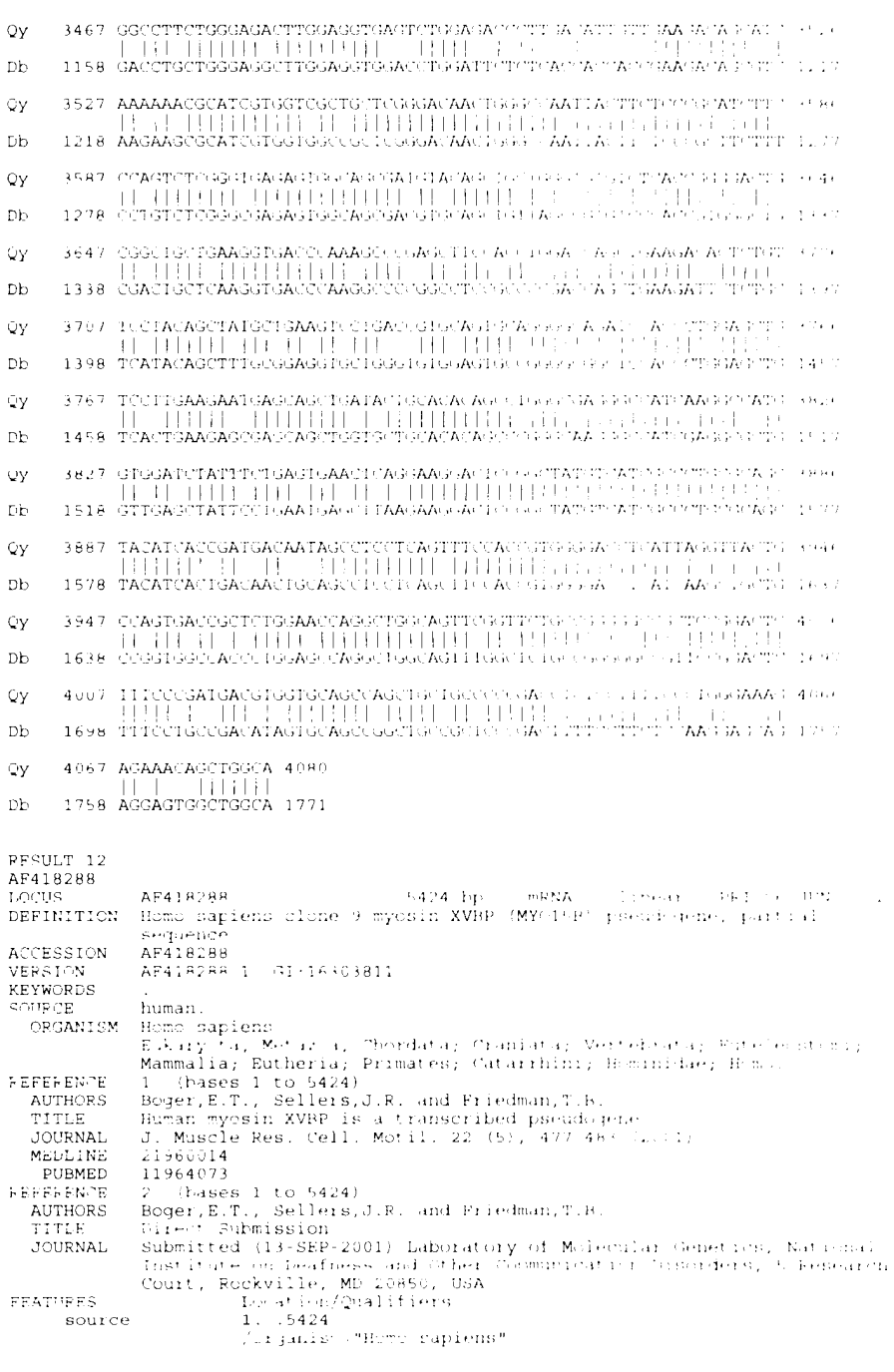




RESULT 11
AX253082
LOCUS AX253082 4174 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168866.
ACCESSION AX253082
VERSION AX253082.1 GI:16986279
KEYWORDS .
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4174)
AUTHORS Brooks,A.R., Deng,G.G. and Pubanyi,G.M.
TITLE Estrogen-regulated unconventional myosin-related protein: compositions and methods of use
JOURNAL Patent: WO 0168866-A 5 20-SEP-2001;
SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1..4174
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Partial DNA sequence for human myosin related protein variant 1 (hMRP1)"
BASE COUNT 901 a 1339 c 1237 g 697 t
ORIGIN







RESULT 12
AF418288
LOCUS AF418288 5424 bp mRNA linear PRI 10-JAN-2003
DEFINITION Homo sapiens clone 9 myosin XVBP (MYO18B) pseudogene, partial sequence.
ACCESSION AF418288
VERSION AF418288.1 GI:16803811
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5424)
AUTHORS Boger,E.T., Sellers,J.R. and Friedman,T.B.
TITLE Human myosin XVBP is a transcribed pseudogene
JOURNAL J. Muscle Res. Cell. Motil. 22 (5), 477-485 (2001)
MEDLINE 21960014
PUBMED 11964073
REFERENCE 2 (bases 1 to 5424)
AUTHORS Boger,E.T., Sellers,J.R. and Friedman,T.B.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, 5 Research Court, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1..5424
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="9"
                     /tissue_type="kidney"
     gene            <1..>5424
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
ASE  UNIT       2.6 ,   1230 ;   1010 3    796 t
RIGIN

 Query Match              19.7%;  Score 861;  DB 9,  Length 5424;
 Best Local Similarity    69.8%;  Pred. No  4.8e-186;
 Matches 1406;  Conservative    0;  Mismatches  460;  Indels  143;  Gaps   17;

Qy       8 GGTGGGAGTTTTCAGCTTACCAGGTGGCAGAATTTATAAACAGAAACAGGGGCTAGCTGGAG 67
          || |||||||||||||||||||| |||||||  |||||||||| ||| || |||||||
Db    3560 GCAGGGACTTGTTACTTACTAGTTTCAAAATTTTTAAACAGAAACCGGGACTAGCTGGAG 3619

Qy      68 CAGGTAGTGTTGGAGATGCTGCAGCCAGAGCCAGTTGCAGTGACCTAGCCTTCTTTTGAG 127
           ||||||| |||||||| | |||||||||| ||| ||
Db    3620 CCTGCTGTTGTTGAGATGCTTGGCTAGAGCAGTTCCAGCTGGC- ---- - - - - - 3663

Qy     128 CTCATGGGCAGTTTGTTTAAGAAGAAGAGCTCAGTCAGTGGAGTGAGCAAAATAAACCC 187
           |||||||||||||||| || |||||||||||| | ||  | || | || ||||
Db    3664 -----GGGCAGTCTGTTTCAGAAGTAGAAGCTCAGTCAGTGGAGTGGGGAGGCAGACCC 3708

Qy     188 ACATTGGCCTCTCGATTCCAGCAGACTCTGGTTGACTTGCTAGCTGGCTAGCAGCAGG 247
           ||  |||||||||| |||||||||||||||| |||  |||| |||||||||| ||||
Db    3709 ACGCTGGTTTTTGGTTTCAGAGGCTTTGAGGAGCCTCATAGCCGGCTGG---GCAGG 3765

Qy     248 GGGCATTTCTAGTTCAGCAGTTTTTCAATTCCAGCTGTGAAAGATGCAGGCCTTTTG 307
           |||||||||| |||| || |||| | |||||||||| |||||||||||||||||||
Db    3766 AGCCATGTCTATTTCATCCAGTGCCTCAGCCTTAACCTTGGAAAGCTTCCAGGCCTTTT 3825

Qy     308 GACGTGGGGCATGTGGCAGAGCAGTTGTCTAGCTGTGCATCCTGGAGATCATAGGCAGT 367
           |||||||| |||||| |||||||||| |||||||||||||||| |||| |||||||||
Db    3826 GACGTGGGACATGTGACAGAGCAACTGTACAGCCAGCCATATCTGGAGGCCCTGGCAGCC 3885

Qy     368 GGGATATTTTAGTTTTTTTGGAATGTCCTTTTAAGTCTTTCTGGCAAGGTTCCATGCC 427
           || || || |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3886 CGCATTGCCAACTTCCCCCCCC- - - - - - -ATGCCTTGGAGCAGCTTCGGAGGCC 3930

Qy     428 CTGGGGTCAGGGAGACAGAAAGTTGTTTTCAAAGGAAGGTGTGGTGGCCATCCTCACT 487
           |||||||||   ||| |||  |||||||||||||||||||||||||||| ||||||||
Db    3931 CTGGTGTCAGAAGGCAGGAGACTTTTGACGGGGAGAAGTGTGGTGGCCCTTGAGCTGAG 3990

Qy     488 GAAGTTGTCCCCCCAGAGTCAGTCCTGTATCATCTTGGAGTCACCCAGGTCCTGGTGTAG 547
           |||||||||| || || |||||||||||||||||||||||||||||||||||||||||
Db    3991 GAGGTGCTGGTGGCAACTACTTTCTATACATCTTGGAGCCACCAAGGTCCTGCTGCAG 4050

Qy     548 GAACAAGGCTGGCAGGAGCTAGAACAGCTGTGGGTTCAGGGTGGTCACAGTCCCTGTTC 607
           |||||||||||| ||| || |||| ||| |||| |||||||||||||||||||||||
Db    4051 GAGCAGGGCTGCCAGCCGCTTGAGGAGCTCCCGGACCAGCAGCCCTCCCAGCCCTGGTC 4110

Qy     608 ACTCTTTAGCCCTGCCTCCCAGCCCTGTATCAGCCGGCAGCGGCTCCGTCTCCTGCCCCGG 667
           |||||||| || ||| |||| |||| |||| |||||||||||||||||||||||||||
Db    4111 GACCTGCAGCGGCAGCTTCCACACCTGCATCTTCGGGCAGGGC ---- GTCCTGCCCGG 4164

Qy     668 ATGCAGGGTCCCTGTGTTCCTGGCTTCCAGGCAGAGAAGCGATATCTCCAGCGGAGCTCAGCT 727
           |||||||||| |||||||||||||||||| |||||||| |||| || |||| |||||||
Db    4165 ATGCAGGCTCGCATGCGTGGGTTCCAGGCA--GAAGCGGTACCTCCGGCGGCGGGCAGCT 4223

Qy     728 CTGGAACAGCTGAACACCATTCTCCTATGTGGCTCGGCCCCTGCTCCGGAGACGACAGAAG 787
           |||||||||||||||| || |||||||||||| |||||||||||||||||||||||||
Db    4224 TTGGGACAGCTGAACAGCATCTTGTGTGGGCCAGCTGGTCCGAGAGGCGGCAGAGA 4283

Qy     788 CTACGGTGTGGCCCTGGCCC- ----  -CACAGCGGGGAGCCCTGGGGGAAAGTGTCA 832
           |||| || |||| |||| |||  |||||  | ||| || || |||| ||
Db    4284 CTGGAGTTTGGCCTTGCAGGCCCGGCATAGTAGCGAAAGGGCCTTGGAGAGATTGCCA 4343

Qy     833 AATATGGACCTGGGTCGCTTAGAGATCCCCGCCAGCTGGCTACTCTGCTGGAGAGGGCG 892
```

```
Db    4344 AGCATGGAGCTGGCGCCTTGGAGATCCCGGCTGAGCTGGCCGTCATGCTGAAGACGGCG 4403

Qy     893 GAAGGCCACCAGGCCTTGGTGACGGGAGCATCACAGAGTCCCTGCCACCTGAGGTCCCC 952
           ||| ||| | || |  ||| | |||||||||| |||| |||||| ||||||| ||
Db    4404 GAAAGCCATCGGGACCGCCTGGCTGGGAGCATCACCGAGTGCCTGCCGCCTGAGGTTCCT 4463

Qy     953 GCCCGGCCCAGCCTGACTCTCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTA 1012
           |||||||||||||||||||||||||  |||||||||||| || || |||||| || ||
Db    4464 GCCCGGCCCAGCCTGACTCTCCCAGCAGACATTGACCTGTTCCCTTTCTCCAGCTTGGTC 4523

Qy    1013 TCCACCAGCTTTCAGAAGCCATTTCTGGCTCAACAGTCAACTCAGGAGAGCCCCTA 1067
           ||| | ||||||||                ||||||||||||||| | ||||||
Db    4524 GCCATCGGCTTTCAGG-- - - - - - - - CCAGGGCAGCCACT GCGAAGCCCCTA 4565

Qy    1073 ACGCGGTTAGATGGCGAGAACCCTCAGCAGGCTCTGGAGATCAACAAGTGATGCTGCGG 1132
           || | | |||| || |||||||||| || |||||||||||||||||||||||||||
Db    4566 ACCCAGCTGGATGGAGACAACCCTCAGGGTGCCTGGACATCAACAAAGTGATGCTGCGG 4625

Qy    1133 CTCCTGGGGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCACGTTCCTCGTG 1192
           |||||||||| |||| ||| |||||||||| |||||||||| |||| |||| | | |||
Db    4626 CTCCTGGGGGACGGATCCCTGGAGTCCTGGCAGAGGCAGATGATGGGCCCATACCTGGTG 4685

Qy    1193 CAGCAGGCCCAGCGAGGGCGGGACTCCGAGATGAGCTCTTCAGCCAGCTCGTGGCCCAG 1252
           | |||| ||| | |||| |||||||||||||||||||||||||||||||||||||||
Db    4686 CGGCAGGGGCAGTGCGGCCAGGGCTGTGAATGAGCTCTTTAGCCAGCTCGTGGCCCAG 4745

Qy    1253 CTGTGGCGCAAGCCAGATGAGCAACAGAATCAGGGTGGCTGGGCCCTAATGCTGATCCTG 1312
           || |||||||||| || |||| |||| |||| |||||||||||||||||||||||||||
Db    4746 CTATGGCAGAAGCCAGATGAACAGCAGAGCAGTGGTGGCTGGGCCCTCATGGCTGTTTTG 4805

Qy    1313 CTCAGCTCGTTTGGTCGCACACTCTGCCTGGAGAAGCCACTGCTGAAATTTGTATCTGAC 1372
           |||||| |||| | ||  |||| |||  |||||||||| |||||||||||||||||||
Db    4806 CTCAGCGCCTTTGGCCGCATGCTGTTCCAGAAGCCACTGCTCAAGTTCGTCTCTGAC 4865

Qy    1373 CAGGCTGCCAGTGGCAGCACAGCTCTGTCAGCAGAAGCCGCTGCAGGTGCCCTGGAGCAG 1432
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4866 CAGGCTGCCAGTGGCATCACAGCTCTGTCAGCAGAAGCCGCTGCAGGTGCCCTGGAGCAG 4925

Qy    1433 ACAGGGTGGGCTCCTATGGCTTCAGGTCTACCAGCCACCACAACTTGAGTGGAAGGCT 1492
           | | |||||| | |  |||| |||||||||||||||||||||||||||||||||||||
Db    4926 TCGTAGCTGGTCTTCTGGGCAGTCGTGGCCTACCCCCAGCCAGCTGGAGTGGCTGGCC 4984

Qy    1493 GGTTTACGTCGGGCCGCATGGCGCTGGAGTGTTTCACATTCAACGAGGAAAGCTACTCC 1552
           |||  | | |||  |||| |||| |||||||||||||||||||||||||||||||||||
Db    4985 GG- ---ATGGCGGGTCAGCCAGTGGATGGAGAACACCCCAGCCGTGCCCTGGACATC 5039

Qy    1553 GGGGAAGTGTGAATGTTGGACCACGGGAGAACATGTTTGACGCCGATCCTACAGAGCAGA 1612
           |||||  || | || | |  |||| |||||||||||||||||||||||||||| |||
Db    5040 AACAAAGTGATGCTGCGGCTCCTGGGGGAGGGATCCCTGGAGTCCTGGCAGAGCCAGATC 5099

Qy    1613 GGGCTGGAGGGCGCCCTGTGGCTGGTCTGTCTGTCACTCCATCGTGGGGATGCTTGCCT 1672
           |||| | || | |||| |||||||||| |||| |||||||||||||||||||||||||||
Db    5100 - ---ATGGGCGATACCTGGTGGGGCAGGGCTGTCACTCCATCGTGGGGATGCTTGCCT 5145

Qy    1673 GACTTGCTTGGCTGTGACTTTGGTGTTGAGCTAATAGCCAGACTGAGGACTTGGGAGAC 1732
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5146 ----------------------------------------------------GGAA 5149

Qy    1733 GCAGCTGGTCCCCACAACTACCCCATCACTCCTCTTGGTTTAGCTGAGAGCATCCCTCCA 1792
           |||| | |||| || |||||||||||||||||||||||||||||||||||||||||||
Db    5150 TGAGTTGGCCCCGCCAGCTACCCCATCACTCCTCTTGGCTTAGCCGAGGCCATTCCCTTT 5209

Qy    1793 GCCCCTGGTGTCCAGGCTCCTTCCCCGCCCCCAGGACCCCCTCCAGGTCCAGCCCCAACG 1852
           |||||||| ||||| || ||||||||||||||||||||||||||||||||||||||||||
Db    5210 GCCCCTGGCATTCAGGCCCCTACACCCCCCAGGACCCCCTCCAGGTCCAGCCCCAACG 5269

Qy    1853 CTGGCCAGCAGCCGCCTCCAGAAGAGTCTAGTAAGCCTGAGAACCTGGATGGTTTCGTG 1912
           ||| |||||| ||  | || |||| | | | |||||||||||||||||||||||||||
Db    5270 CTGCCCAGTAGGGACCACACAGGGAGTTCCAGAGCCCAGGAAGCCTGGACGGCTTCCTG 5329

Qy    1913 GACCACCTCTTTGAACCAGTCCTCCTTGGGTTTCAGTGATCTGGAACAAGGCTGGGCC 1972
           |||||  |||| | |||| | |  | || |||| |||||||||||| |||||||
```

```
Db    5330 GACTCACATCTTCCAGCCAGTGATATCTCTGGGCCTCAGGGATCTGGAACAAAGCTTGGGCT 5389

Qy    1973 CTTGATCAGATGCATGAAGGGAGGGGGCTCTGTTGG 2007
           ||||||| |||||||||||||||||| |  ||||
Db    5390 CTTGATGAGCCGCATGAAGGGAGGGGGCGCCATTGG 5424


RESULT 13
AF418291
LOCUS       AF418291                4802 bp    mRNA    linear   PRI 06-JUN-2001
DEFINITION  Homo sapiens clone 17 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418291
VERSION     AF418291.1  GI:16303814
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4802)
  AUTHORS   Bauer,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21960014
    PUBMED  11964073
REFERENCE   2  (bases 1 to 4802)
  AUTHORS   Bauer,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..4802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="17"
                     /tissue_type="stomach"
     gene            <1..>4802
                     /gene="MYO15B"
                     /note "myosin XVBP"
                     /pseudo
BASE COUNT      952 a   1508 c   1749 g    693 t
ORIGIN

  Query Match             19.2%;  Score 577;  DB 9;  Length 4802;
  Best Local Similarity   77.6%;  Pred. No. 4.5e-151;
  Matches  707;  Conservative    0;  Mismatches  206;  Indels   10;  Gaps    2;

Qy     698 TCAGGGCAGGAAGCGATATCTCCAGCGGAAGGTCAGCTCTGGGACAGCTGAACACCATTCT 744
           |  |  |||||||| || |||| |||| || ||||| ||||||||||||||||||||| ||
Db    3752 CCTGGAAAGGAAGCGGTACCTCCGGCTGCGGCACAGTTTGGGACAGCTGAACACCATCCT 3811

Qy     745 CCTAGTGGGCCCGGCCCCTGCTCCTCGGAGACGACAGAAGCTACGGTCCTCGCCCTGGCCCG- 802
           | ||.|||| |||||||||| ||| || |||| || | |  || | | | | | |
Db    3812 GCTTGGTGGCCCAGCCCCTGCTC AGAGGCCCACAGACTGCAGCTTGGGGCGTTGGCAGG 3871

Qy     803          CACAGCCGGGAGCCCTGGGGCAAAGTGTCAAATATGGACCTGGGTCGCTTAG 854
                    |||||    | ||| || || |||| |||  ||||| |||| |||| |
Db    3872 CCTCGTTTACCAGCGAAAGGGTCCTTGGACAGAGTGCCAAGCATGGAGCTGGGGCGCTAGG 3931

Qy     855 AGATCCCCCCCCCACCTGGCTACTCTGCTGGAGAGGCCCAAGCCACCAGGCCTTGCTGA 914
           | ||  || ||  |||||||     ||||| ||| ||||||| |||| | || |   |||
Db    3932 ACATTCCCCCCTCACATTGGCCTCATGCTAAATATCCCGAAGCCATCGGCACCGCCTGG 3991

Qy     915 CCCCAAGATCACAGAGTCCCTGCCACCTGAGGTCCCCGCCCGGCCCAGCCTGACTCTCC 974
           |||| |||| |||| |||| |||||| || ||||||||||||||||||||||||||||
Db    3992 CCCGAGCATCACAGAGTCCCTGCCGCCTGAGGTCCCTGCCGGGCCCAGCCTGACTCTCC 4051

Qy     975 CTCCATCATTGACCAGTTCCCCTCCCAGCTCGATCCACCAGAAGCCCAT 1024
           | |||||| ||| || |||| ||| || ||  ||| | |||||||| | || |
Db    4052 CACCATCATTGACCTGTTCCCTTTCCCAGCTCGTCCACCATGGCTTTCAGAACCCGT 4111

Qy    1025 TTCTGCCCTGGACCAGGGCAGCCACTGGACGAGCCCCTGACCCCTTACATGCCGAGAACT 1084
           |||||  | ||||||||||||||  |||||  |||| |||| |||| |||| |||| |
Db    4112 CCCTGCCCAGGCCAGGCCAGCCACTCCGCAAACCCCTAAC CCCTCACATCACAACT 4171

Qy    1085 CTCAGCAGGCTCTGGAGATCAACAGGGTGATCGTGCCCCTCCAGGAAGGATCTCCTCT 1144
           ||||||  || |||| ||||||  |||||||||||||||||||||||||||||||||||
Db    4172 CTCAGCGTGCCCTGGACATCAACAAAGTGATCGTCCCCCTCCAGGATCCTCCTCT 4231

Qy    1145 AGTCCTGGCAGAGGCAGATCATGGGCGCATACCTGCCCCCCAGCCCAGGATGCTCCCCT 1204
           |||||||||  |||| |||||| | | || ||  |  |||| | |||| |||| |||| |
Db    4232 AGTCCTGGCAGAGGCAGATCATGGGCGCATACCTGCCCCCCAGCCCAGGATGCTCCCCT 4291

Qy    1205 GACTCCGGAGATGAGTTTTCAGCCAGTTGGGGCCCCAGCCCCCCAGCCAGCAGCCCCCC 1264
           |||| ||  |||||||||| ||||||||||||||||||||||| | |||| |||| |||| |
Db    4292 GGCTGCGGAAATGAGCTTTTAGCCAGCTGGTGGCCCCAGCTATGCCAGAACCCAGATGAAT 4351

Qy    1265 AACACAAATACCCCCCCCCCCCCCCCCCATCCTCACCCCCCCCCCCCCCCCCCCCCCCCC 1324
           | ||||  |||||||||||||||||| ||||  | |||||||| |||| | ||  |||| |
Db    4352 AGCAGAGCCAGCGTGGCTGGGCCCTCATGGCTGTTTGCTTAGTGCTTTTCCCCCACTGC 4411

Qy    1325 CTGCCCTGCAGAAGCCACTGCTCAAGTTCGTGTGTGCAGCCAGCCCCCCCCCCCCCCCCC 1384
           ||| |||  |||||||||||||||| || || |||||||| || |||| ||||||||||
Db    4412 CTGTGCTACAGAAGCCACTGCTCAAGTTCGTGTGTGCAGCCAGCCCCCCCCCCCCCCCCC 4471

Qy    1385 CCCTGTGCCAGCACAAGCTGCTGAGGGCTTGGAGCAGTACTCTTACTCTTCCCCCCCCCC 1444
           | |||||||||||||||| | || ||| |||||||||||||||||||||||||||||||
Db    4472 CGCTGTGCCAGCACAAGCTGCTGAGGGCTTGGAGCAGTACTCTTACTCTTCCCCCCCCCC 4531

Qy    1445 CGAGGTCCCACCCACCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1504
           |  || |||||| || || || || |||||  ||||||||||||||||||||||||||||
Db    4532 CTCGGCCCACCCCCCGACCCAGCTGGAGTGGCTGGCGGATCTCTTCTGGGGCTCTATCC 4591

Qy    1505 CGCTGGATGTGTTCACATTCAAGGAGGAAAGCTACCCCCCCCCCCCCCCCCCCCCCCCCC 1564
           |||||||||||||||  ||| ||||||  |||||||| |  |  |||| || || ||||
Db    4592 CGCTGGATGTGTTCAGCTTCAGGGAGGAGTGCTACCTGGCCATCTTCAGCTTTGGACCA 4651

Qy    1565 CGGGAGAGCAGTTTGCAGGGTGGATCCTACAGAGCCAG 1611
           |  || |||||| | || |||||||||| |||||||
Db    4652 CGGGGGAGCAGCTGGCTGGGTGGATCCTGCAGAGCAG 4688


RESULT 14
AK025259
LOCUS       AK025259                2099 bp    mRNA    linear   PRI 06-JUL-2001
DEFINITION  Homo sapiens cDNA: FLJ21606 fis, clone COL07422.
ACCESSION   AK025259
VERSION     AK025259.1  GI:10437234
KEYWORDS    oligo capping; fis; full insert sequence.
SOURCE      Homo sapiens colon cDNA to mRNA, clone fib07422, clone_lib COL.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inamoto,R., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Kayasuki,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2099)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Okayasu,M., Nishi,T., Inamoto,R.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639,
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5300,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5' & 3' end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
```

University of Tokyo (partly supported by Science and Technology Agency).

```
FEATURES             Location/Qualifiers
     source          1. .2099
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CO[illegible]302"
                     /tissue_type="colon"
                     /clone_lib="COL"
                     /note="cloning vector pME18SFL3"
BASE COUNT       433 a    [illegible] c    606 g    370 t
ORIGIN
```

```
 Query Match             12.4%;  Score 544.2;  DB 9;  Length 2099;
 Best Local Similarity   80.6%;  Pred. No. 1.5e-113;
 Matches  [illegible];  Conservative    [illegible];  Mismatches  153;  Indels    0;  Gaps    0;

Qy   3281 TCTGGAAGTGTTGTTCTTCTTATCCAATTACCCTGGAAGTTGTTCTTACCCAAGGAG 3340
          |||  ||  |||| |  |  |||  ||  ||||||||||||| ||||||||||
Db    922 TCTTCCAGCCAGTGATATCCTCCTTCTTAGTGATCTTGGAAGTTGTTCCTACCCAAGGAG 981

Qy   3341 GTGTTCTACCCAGGGGAGAACTTCAGTTACCCATACTGCCTGAGTCTCCTCTGTGCCAGCAG 3400
          ||||||||||| |||||||||||||| ||||| |||| ||| || |||||||  |||||
Db    982 GTGTTCTACCCAGGGGAGAACTTCAGTTATCCCTACTACCTGAGGCTCCTCTGTGACCAG 1041

Qy   3401 ATCCTGGGGGACACCTTCACAGAGTCCTGCACCGGGATCTCACAGGATGAGCGGCACAAA 3460
          |||||  |||||||||||||  |||||||  | |||||  || |||||||||  |||
Db    1042 ATCCTGAGGGACAGCTTCATCCAGATCCCTACTTTCGCAGAATGAGCGGCGGCAAA 1101

Qy   3461 ATGAAAGGGGTTTTTGGGACTTGGAGTGAGTCTGGAGACCCTTGACATTGTTGAAGAC 3520
          |||||||  |||||||||||||||||||||||  |||||  | ||  ||  |||||
Db    1102 ATGAAAGACTTTTTTGGGAGTTTGGAGTGGACCTGGATTCTCTCACCACCACGGAAGAC 1161

Qy   3521 AGCATCAAAAAACGCCATCGTGGTCGTCTTCGGAACAACTGGGCCAATTACTTCTCCCGC 3580
          |||  |||| ||  |||||||||||  ||  |||||||||||||||||||||||||||||
Db    1162 AGCGTCAAGAAGCGCATCGTGGTGGTCTTCGGAACAACTGGGCCAATTACTTCTCCCGC 1221

Qy   3581 ATCTTCCCAGTCTCGGGCTGAGATTGTACCATTGTACAGCTGCTGGGTGTGTCTCACCGG 3640
          |||| || |||||||||||||||||||||||||||||||||| | |  |||||  |||||
Db    1222 TTCTTTCCTGTCTTCGGGCTGAGAGTGGCACCAGAACTGGCAGCTGTTACCGGTGTCCCACCGT 1281

Qy   3641 GGACTGGCACTGACTGAAGGTGACCTAAAGCTCGAGCTTCCACCTGGACCAGCTGAAGACA 3700
          ||  |||||  |||| ||||||||||||  ||||  || ||  ||  ||||||||||||
Db    1282 GGGCTGGGATTGCTCAAGGTGACCCAAGGCCCAGGCTCGGCCCGACCAGCTGAAGATT 1341

Qy   3701 CTCTGTTTCCTACAGCTATGCTGAAGTCCTGACCGTGCAGTGCAGGGGCAGATCCACCCTG 3760
          |||||  || |||||||  ||| ||  |||  ||| |||||  | |||||||||||||||
Db    1342 CTCTGATTCATACAGTTTTCTGGAGGTGCTGGGTGTGGAGTGCCGGGGCGGCTCCACCCTG 1401

Qy   3761 GAGCTCTTTTTTAAGGAATGAGCAGCTGATACTGCACACAGCCTGGGCGAGGGCCATCAAG 3820
          |||||||| ||||||  |||||||||| | |||||||||||| |||| |||||||| ||
Db    1402 GAGCTGTTATTGAAGAGCTAACAGCTGGTATTGTACACAGCCGGGGCAAGGGCCATCGAG 1461

Qy   3821 GCCATGGTGGATCTATTTCTGAGTGAACTCAAGAAGGACTCCGGCTATGTCATCGCCCTG 3880
          ||  |||| || |||| |||| |||  | | ||||||||||||||||||||||||||||
Db    1462 GCGCTGATTAAGCTATTCCTGAATGAGTTTAAGAAGGATCTCGGCTATGTCATCGCCCTG 1521

Qy   3881 TCCAGCTACATCACCTATGACTACGAATACAGCTTCTACTTTCCACCGTGGGGACCTCATTAGC 3940
          |||||||||||| ||  |||||||||||||||||||||||||||||||||||| | |
Db    1522 CCCAGCTACATTACTGACAAGTCCAGCCCTCCTTACTTCCACCGTGGGGACCTCATCAAG 1581

Qy   3941 TTACTGCCAGTGACCGCTCTGGAACAGGCTGGCAGTTCGGTTCTGCCGGGGGCCGCTCC 4000
          | |||||  ||| ||  | |||| ||||||||||||| || ||||||||||| |||
Db    1582 CTGCTGCCGGTTGCCAAGTTCGAGCAGGCTGCAGTTTGGCTCCTGCCGGGGGCCGTTCC 1641

Qy   4001 GGACTCTTTTCCCATGACCTCGTCACCGCAGCTGTTGCCCGGACCTCTCCTTTTTCCTTG 4060
          |||||||||  |||| |||||||||| |||||||||| ||||||||||||  |||| ||
Db    1642 GGACTGTTTTCCTGGCCACATACTGCAGCCGCCTGCCCGACTTTTCCTTCTTCTTG 1701

Qy   4061 GGAAAGAGA 4069
          |  |  ||
Db    1702 AGCAGAGGA 1710
```

```
RESULT 15
AL645647
LOCUS       AL645647              147804 bp    DNA     linear   ROD 30-APR 2002
DEFINITION  Mouse DNA sequence from clone RP23-332L17 on chromosome 11,
            complete sequence.
ACCESSION   AL645647
VERSION     AL645647.8  GI:20386856
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 147804)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 2, 2002 this sequence version replaced gi:20268853.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a copy of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone, and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Note: 332L17 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            See further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .147804
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-332L17"
                     /clone_lib="RPCI-23"
BASE COUNT    34847 a  38451 c  38280 g  36226 t
ORIGIN
```

```
 Query Match             10.6%;  Score 468;  DB 10;  Length 147804;
 Best Local Similarity   69.8%;  Pred. No. 6.8e-94;
 Matches  977;  Conservative    0;  Mismatches    0;  Indels  379;  Gaps    4;

Qy   3476 GGAGACTTGGAGGTGAGTCTGGAGACCTTGACATTGTTGAAGACAGCATCAAAAAACGC 3535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  48411 GGAGACTTGGAGGTGAGTCTGGAGACCTTGACATTGTTGAAGACAGCATCAAAAAACGC 48470

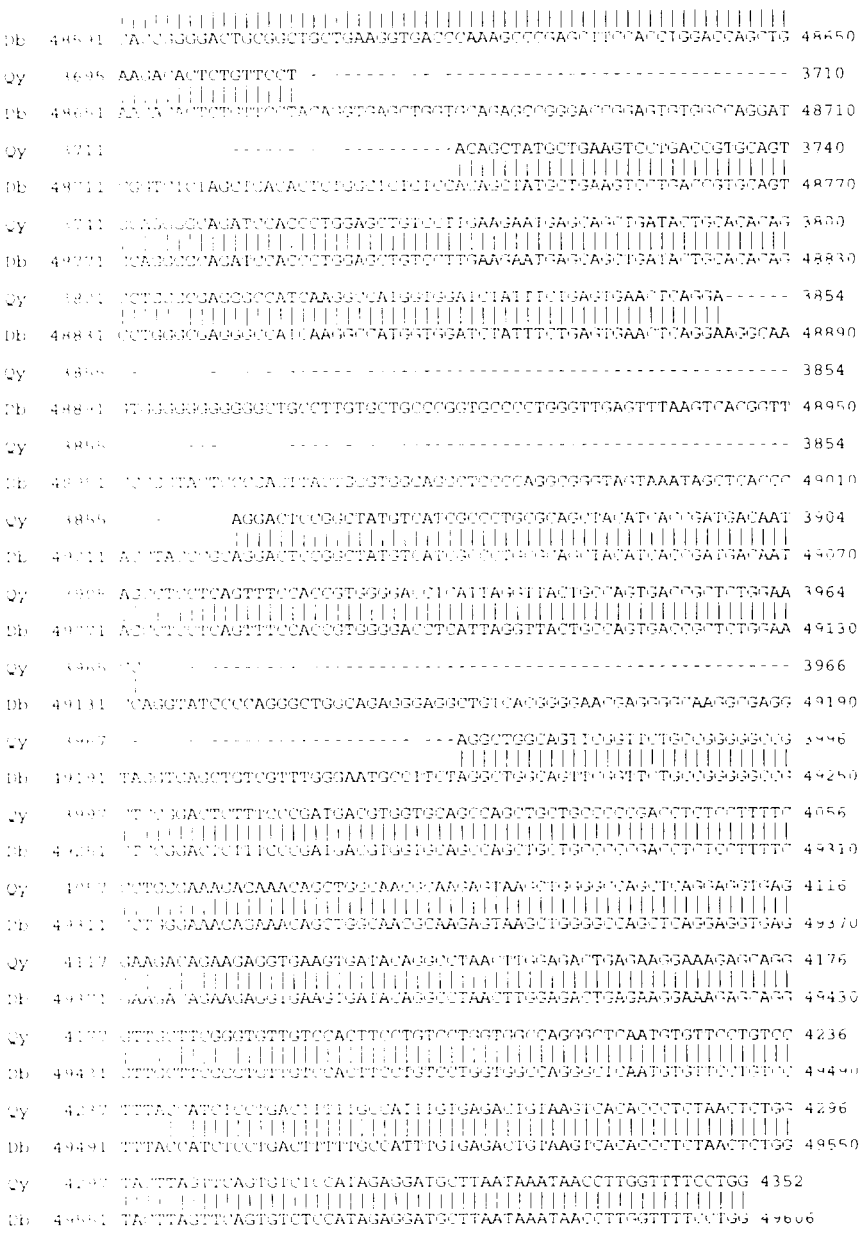
Qy   3536 ATCGTGGTCGCTGCTCTGGGACAACTCTGCCAATTACTTCTCCCGCATCTTCCCAGTCTCG 3594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  48471 ATCGTGGTCGCTGCTCTGGGACAACTGGGCCAATTACTTCTCCCGCATCTTCCCAGTCTCG 48530

Qy   3595 ---------------------------------------------------------- 3594

Db  48531 GTCAGTGGCTGGGTTTAAGGAGTGTGACTCAGCTGCAGCTTTGGGCCGGGAGCGGTGACTG 48590

Qy   3535 ---- -- - -- -- - -GGGTGAGAGTGGCAGCGATGTACAGCTGCTCGGTGTGTCT 3634
                                ||||||||||||||||||||||||||||||||||||
Db  48591 ACCCACTCGGCTTCTCGACAGGGTGAGAGTGGCAGCGATGTACAGCTGCTCGGTGTGTCT 48590

Qy   3635 CACCGGGGACTGGGCTGCTGAAGGTGACCCAAAGCCCGAGCTTCCACCTGGACCAGCTG 3694
```



Search completed: January 9, 2003, 19:55:51
Job time : 11158.2 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

DNA nucleic - nucleic search, using sw model

Run on: January 9, 2003, 22:42:16 ; Search time 761.196 Seconds
(without alignments)
12042.430 Million cell updates/sec

Title: US-09-803-126-3
Perfect score: 4375
Sequence: 1 cgatgggactgtaaatatt aaaaaaaaaaaaaaaaaaa 4375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125990169 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2[illegible]

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4375	100.0	4375	22	AAS13820	cDNA #2 encoding m
2	4352	99.5	6293	22	AAS13819	cDNA #1 encoding m
3	946.4	21.6	3780	22	AAS13823	DNA encoding human
4	946.4	21.6	4174	22	AAS13821	DNA encoding human
5	706.6	16.2	1476	21	AAZ56716	Human transmembran
6	628.8	14.4	1044	22	AAD07791	Human secreted pro
7	626.4	14.3	1268	23	ABK43544	DNA encoding novel
8	621.2	14.2	1247	22	AAS27113	cDNA encoding nove
9	621.2	14.2	1247	23	ABK43853	DNA encoding novel
10	470.2	10.7	826	22	AAD07777	Human secreted pro
c 11	166.2	3.8	24757	22	AAS27687	DNA encoding novel
12	166.2	3.8	24757	22	AAS33481	DNA encoding human
c 13	94.4	2.2	526	22	ABA60390	Human foetal liver
c 14	89.4	2.0	526	22	ABA29621	Probe #7087 for ge
c 15	89.4	2.0	526	22	AAK08667	Human brain expres
c 16	89.4	2.0	526	22	AAK34556	Human bone marrow
c 17	89.4	2.0	526	22	AAI40274	Probe #8960 used t
c 18	89.4	2.0	526	24	ABS00192	Human genome deriv
c 19	75.2	1.7	170	22	ABA72929	Human foetal liver
c 20	75.2	1.7	170	22	ABA38498	Probe #16964 for g
c 21	75.2	1.7	170	22	AAK21359	Human brain expres
c 22	75.2	1.7	170	22	AAK47522	Human bone marrow
c 23	75.2	1.7	170	22	AAI53353	Probe #22039 used
c 24	75.2	1.7	170	24	ABS01604	Human genome deriv
25	64.6	1.5	7794	24	ABN59632	Novel human codin
26	60	1.4	6733	23	ABL07105	Drosophila melanog
c 27	60	1.4	14200	23	ABL07104	Drosophila melanog
28	58.2	1.3	6171	22	AAD19119	Angiogenesis assoc
29	58.2	1.3	9521	22	AAD19120	Angiogenesis assoc
30	54.8	1.3	7275	23	ABL04215	Drosophila melanog
31	54.8	1.3	10225	23	ABL04214	Drosophila melanog
c 32	54.2	1.2	6655	22	AAH99691	Human protein enco
33	52.8	1.2	7516	22	AAK51987	Human polynucleoti
34	52.8	1.2	8496	22	AAK52971	Human polynucleoti
35	52.6	1.2	5574	24	ABK35570	Gene MYH11 differe
c 36	52.6	1.2	6607	22	AAH98321	Rabbit EST-derived
c 37	52.2	1.2	10732	21	AAA10594	Gene encoding a su
38	51	1.2	7736	23	AAS65910	DNA encoding novel
39	50.6	1.2	6390	23	ABL12023	Drosophila melanog
c 40	49.6	1.1	386	22	AAP65986	Novel human polynu
41	49.6	1.1	1388	21	AAC59039	Human secreted pro
c 42	49.6	1.1	1388	24	ABQ54638	Human ovarian anti
43	49.6	1.1	2556	22	AAI59455	Human polynucleoti
44	49.6	1.1	3438	22	AAI61241	Human polynucleoti
45	49.4	1.1	2238	23	AAS74118	DNA encoding novel

ALIGNMENTS

RESULT 1
AAS13820
ID AAS13820 standard; cDNA; 4375 BP.
XX
AC AAS13820;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA #2 encoding mouse myosin-related protein, mMRP.
XX
KW Mouse; oestrogen-regulated unconventional myosin-related protein; MRP;
KW auditory; antiatherosclerotic, osteoarthic, post-transcriptional; neotropic;
KW antiparkinsonian; cytostatic, oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer, cardiovascular disease; deafness; ss;
KW Alzheimer's disease; Parkinson's disease; chromosome 11; immunogenic.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 125..4138
FT /*tag= a
FT /product= "mMRP"
FT /note= "Oestrogen-regulated unconventional myosin-
FT related protein"
XX
PN WO200168866-A2
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P

```
PR   ?? MAR 2001; 2001US-0803126.
XX
PA   SCHE   SCHERING AG.
XX
PI   Brooks AR;  Dent GS;  Kuhanyi CM;
XX
DR   WPI; 2??1-?????/??
DR   P-PSDB; AAU07399.
XX
PT   Novel oestrogen regulated unconventional myosin related protein and
PT   polynucleotide for modulating effects of oestrogen in mammalian cells to
PT   treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS   Claim 4; Fig 3; 108pp; English.
XX
CC   The invention relates to an isolated oestrogen-regulated unconventional
CC   myosin related protein (MRP). Modulating MRP is useful for modulating the
CC   effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC   mammalian cell. When the level of expression or activity of MRP is
CC   increased, the development of atherosclerosis or osteoporosis is
CC   inhibited and when the level is decreased the development of breast
CC   cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC   useful for detecting the presence of oestrogen signaling in a mammalian
CC   cell, by detecting the expression or activity of MRP in the level of MRP
CC   mRNA in the cell. The presence of oestrogen signaling is used to
CC   determine the responsiveness of the cell to the oestrogen or the tissue-
CC   specific distribution of oestrogen signaling in a mammal. MRP is useful
CC   as a marker protein for the investigation of tissue specific and oestrogen
CC   receptor-specific ligands such as oestrogen receptor antagonists and
CC   agonists. Modulators, recombinant forms, derivatives, variants or
CC   fragments of MRP are useful for enhancing or inhibiting oestrogen
CC   signaling in cells and for treatment of breast cancer or other cancers
CC   and for preventing or treating wide variety of oestrogen-related
CC   diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC   disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC   prevention and/or treating hearing loss and deafness. MRP polypeptides
CC   are useful for generating monoclonal and polyclonal antibodies useful for
CC   identifying oestrogen responsive cells, particularly cells responsive to
CC   oestrogen through an oestrogen receptor alpha or beta, to identify
CC   alleles, interspecies homologues and polymorphic variants and in
CC   immunoassays to detect MRP. The present sequence represents the coding
CC   sequence of mouse MRP variant #2 as described in the invention.
XX
SQ   Sequence 4375 BP; 966 A; 1368 C; 1216 G; 825 T; 0 other;

  Query Match          100.0%;  Score 4375;  DB 22;  Length 4375;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4375;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ??????????????????????????????????????????????????????????  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ??????????????????????????????????????????????????????????  60

Qy      61 ??????????????????????????????????????????????????????????  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ??????????????????????????????????????????????????????????  120

Qy     121 ??????????????????????????????????????????????????????????  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ??????????????????????????????????????????????????????????  180

Qy     181 ??????????????????????????????????????????????????????????  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ??????????????????????????????????????????????????????????  240

Qy     241 ??????????????????????????????????????????????????????????  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ??????????????????????????????????????????????????????????  300

Qy     301 ??????????????????????????????????????????????????????????  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ??????????????????????????????????????????????????????????  360

Qy     361 ??????????????????????????????????????????????????????????  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ??????????????????????????????????????????????????????????  420

Qy     421 ??????????????????????????????????????????????????????????  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ??????????????????????????????????????????????????????????  480

Qy     481 ??????????????????????????????????????????????????????????  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ??????????????????????????????????????????????????????????  540

Qy     541 ??????????????????????????????????????????????????????????  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ??????????????????????????????????????????????????????????  600

Qy     601 ??????????????????????????????????????????????????????????  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ??????????????????????????????????????????????????????????  660

Qy     661 ??????????????????????????????????????????????????????????  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ??????????????????????????????????????????????????????????  720

Qy     721 ??????????????????????????????????????????????????????????  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ??????????????????????????????????????????????????????????  780

Qy     781 ??????????????????????????????????????????????????????????  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ??????????????????????????????????????????????????????????  840

Qy     841 ??????????????????????????????????????????????????????????  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ??????????????????????????????????????????????????????????  900

Qy     901 ??????????????????????????????????????????????????????????  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ??????????????????????????????????????????????????????????  960

Qy     961 ??????????????????????????????????????????????????????????  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ??????????????????????????????????????????????????????????  1020

Qy    1021 ??????????????????????????????????????????????????????????  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ??????????????????????????????????????????????????????????  1080

Qy    1081 ??????????????????????????????????????????????????????????  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ??????????????????????????????????????????????????????????  1140

Qy    1141 ??????????????????????????????????????????????????????????  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ??????????????????????????????????????????????????????????  1200

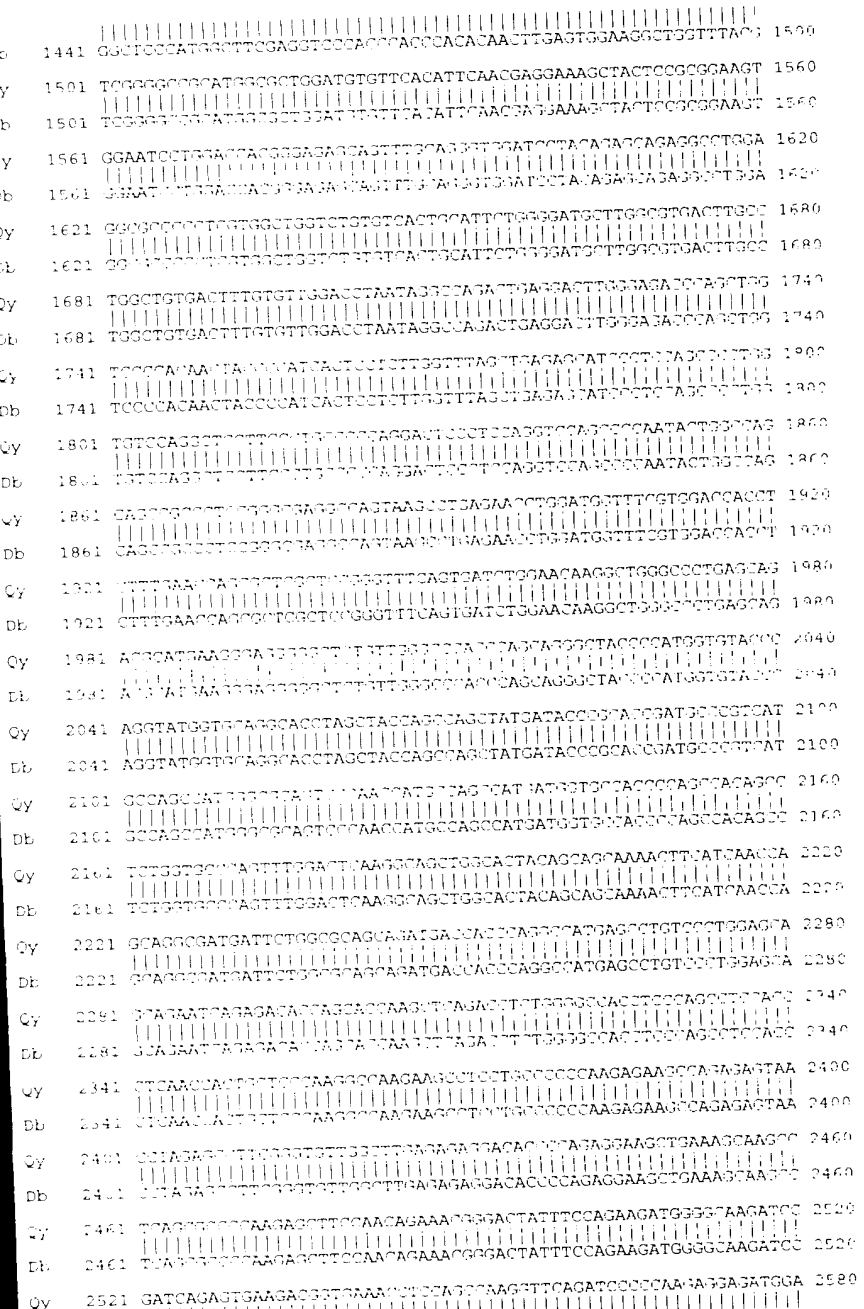
Qy    1201 ??????????????????????????????????????????????????????????  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ??????????????????????????????????????????????????????????  1260

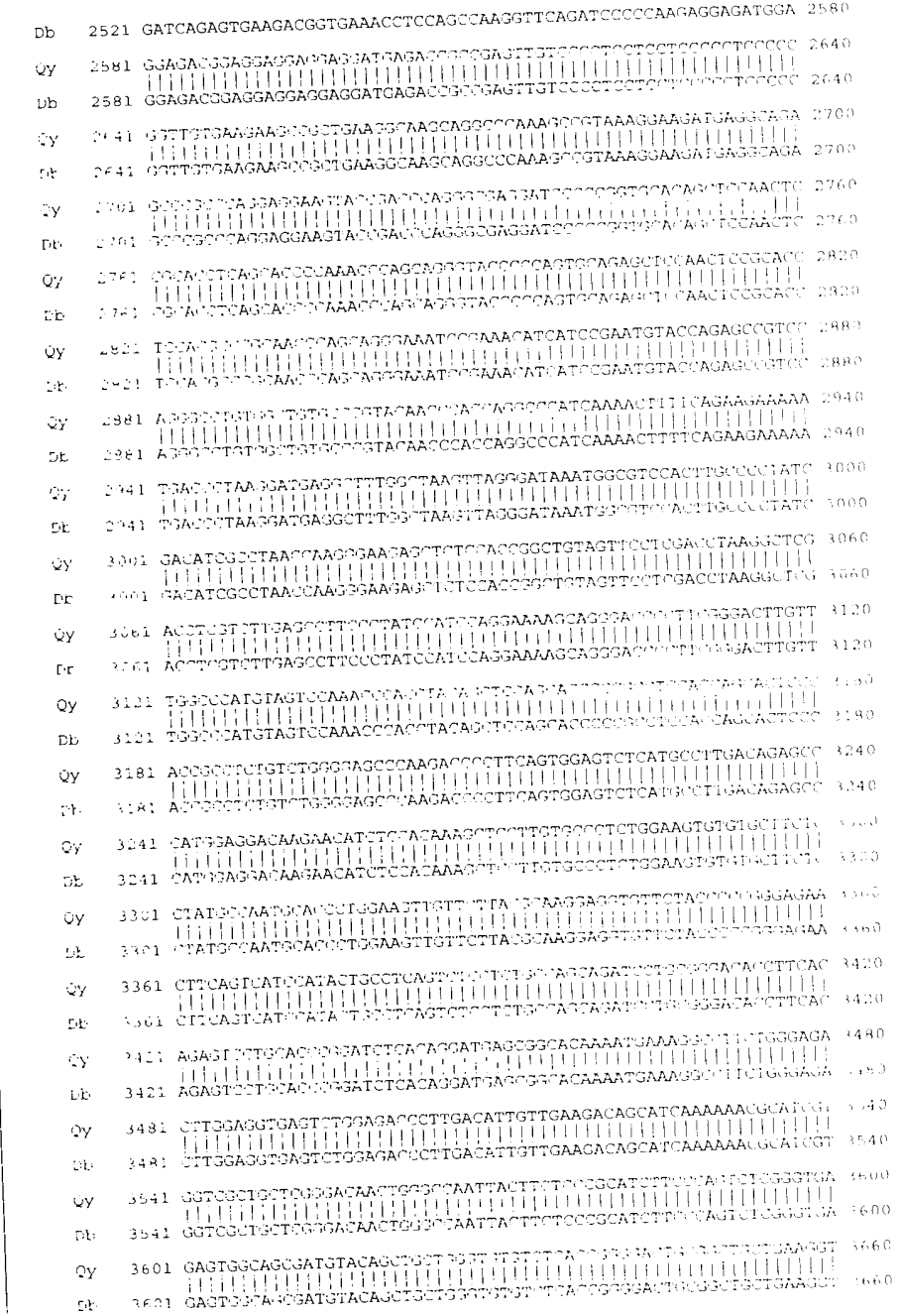
Qy    1261 ??????????????????????????????????????????????????????????  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ??????????????????????????????????????????????????????????  1320

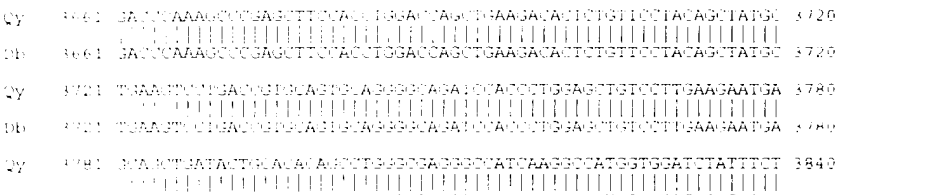
Qy    1321 ??????????????????????????????????????????????????????????  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 ??????????????????????????????????????????????????????????  1380

Qy    1381 ??????????????????????????????????????????????????????????  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 ??????????????????????????????????????????????????????????  1440

Qy    1441 ??????????????????????????????????????????????????????????  1500
```

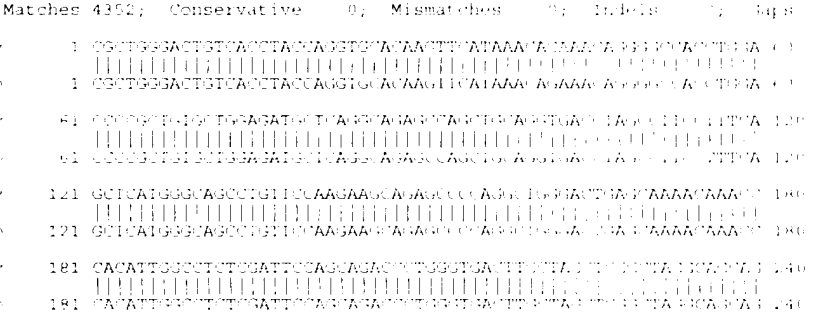


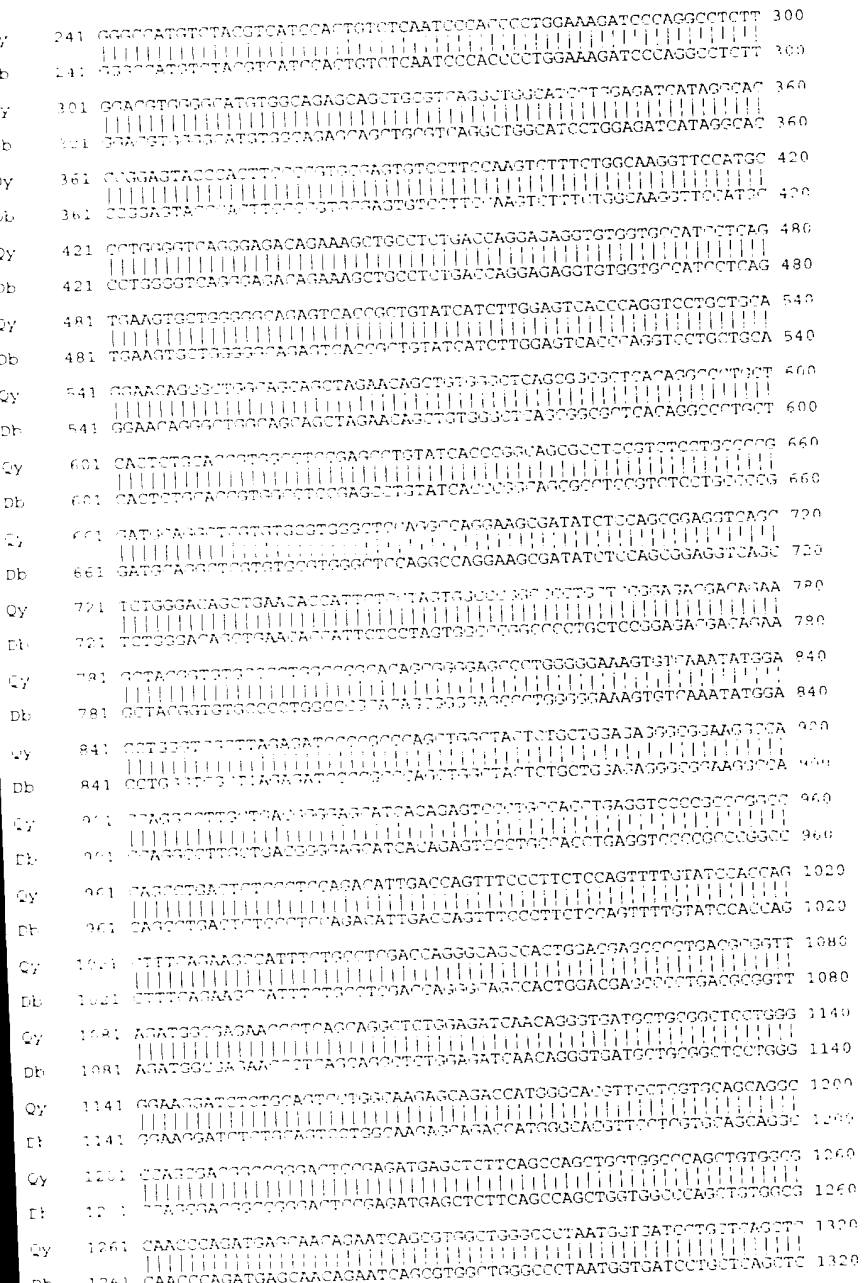


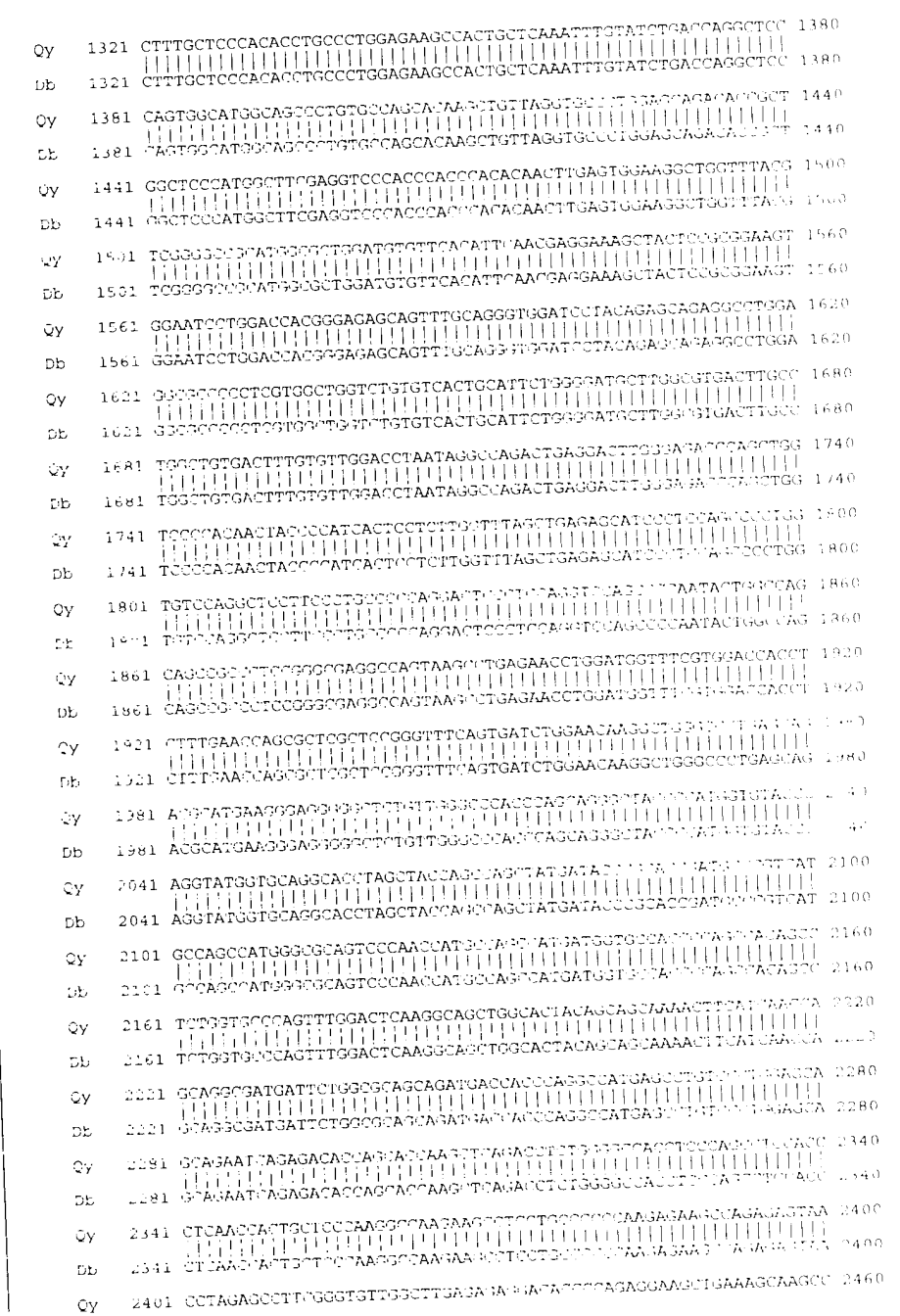


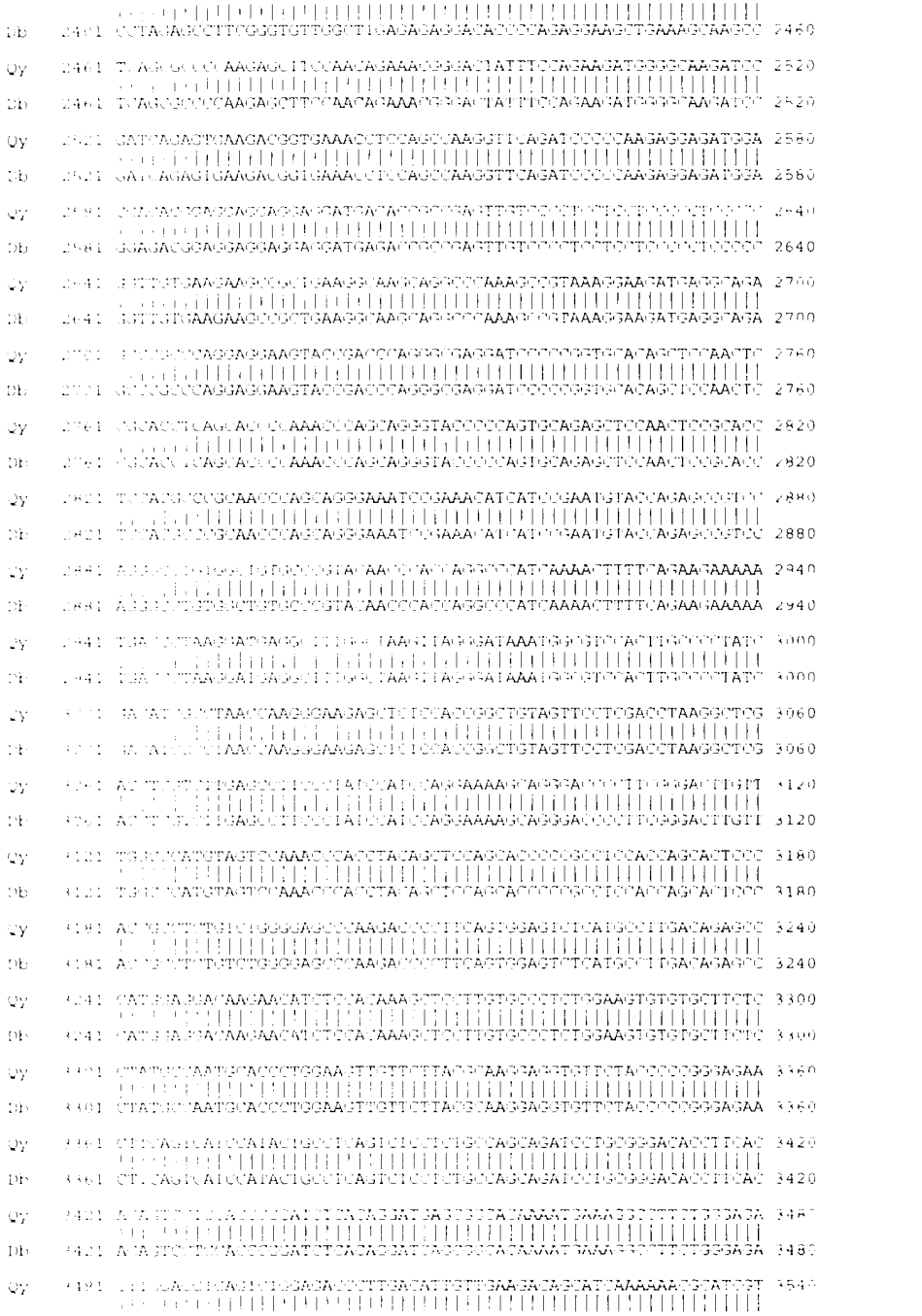
RESULT 2
AAS19819
ID AAS19819 standard; cDNA; 6293 BP.
XX
AC AAS19819;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA #1 encoding mouse myosin related protein, mMRP.
XX
KW Mouse; oestrogen regulated unconventional myosin related protein; MRP;
KW auditory; antiatherosclerotic; osteopathic; neuroprotective; nootropic;
KW antiparkinsonian, cytostatic, oestrogen receptor, atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW Alzheimer's disease, Parkinson's disease; chromosome 11, immunogenic.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 126..4138
FT /*tag= a
FT /product= "mMRP"
FT /note= "Oestrogen regulated unconventional myosin-
FT related protein"
XX
PN WO200168806-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCRI) SCRIPPING AG.
XX
PI Brooks AR, Deng CS, Rohanyi GM;
XX
DR WPI; 2001-602747/68.
DR P-PSDB; AAU07899.
XX
PT Novel oestrogen-regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 4; Fig 2; 108pp; English.
XX
CC The invention relates to an isolated oestrogen regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the murine
CC sequence of mouse MRP variant #1 as described in the invention.
XX
SQ Sequence 6293 BP; 1433 A; 1917 C; 1734 G; 1219 T; 0 other;

Query Match 99.5%; Score 4352; DB 22; Length 6293;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





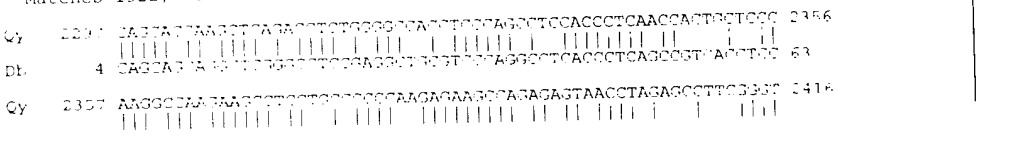


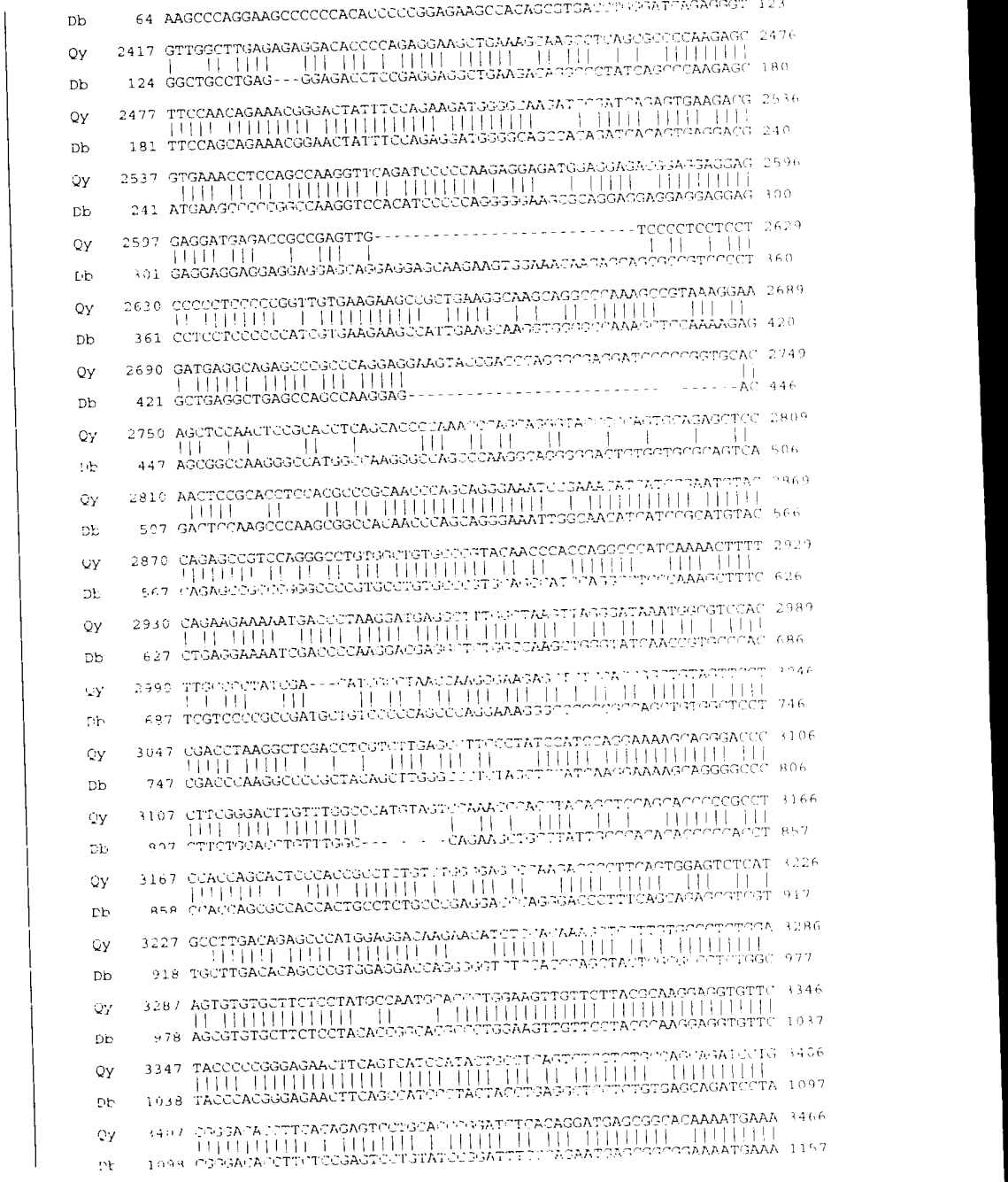


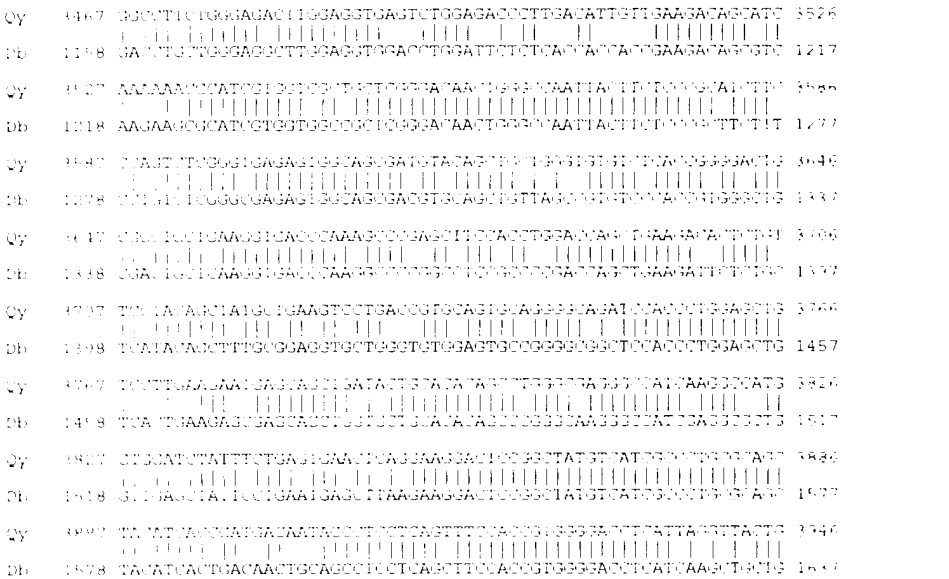
RESULT 3
AAS13823
ID AAS13823 standard; cDNA; 3780 BP.
XX
AC AAS13823;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding human myeloid-related protein, MRP #2.
XX
KW Human; chemotherapeutic; immunosuppressive; neuroprotective; multigrip;
KW auditory; antiatherosclerotic; osteopathic; neuroprotective; multigrip;
KW antiparkinsonian; cytostatic; nootropic; antiepileptic; antiarthritic;
KW osteoporosis, breast cancer; cardiovascular disease; inflammation; MRP;

```
KW   Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             561..2911
FT                   /*tag= a
FT                   /product= "hMRP"
FT                   /note= "Oestrogen regulated unconventional myosin-
FT                   related protein"
XX
PN   WO200168866-A2.
XX
PD   20-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-US08060.
XX
PR   10 MAR 2000; 2000US-188488P.
PR   [illegible] 2001; 2001US-0803126.
XX
PA   (SCHE) SCHERING AG.
XX
PI   Brooks AP, Deng GG, Rubanyi GM;
XX
DR   WPI; 2001-602747/68.
DR   P-PSDB; AAU07406.
XX
PT   Novel oestrogen-regulated unconventional myosin-related protein and
PT   polynucleotide for modulating effects of oestrogen in mammalian cells to
PT   treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS   Claim 4; Fig 7; 108pp, English.
XX
CC   The invention relates to an isolated oestrogen regulated unconventional
CC   myosin related protein (MRP). Modulating MRP is useful for modulating the
CC   effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC   mammalian cell. When the level of expression or activity of MRP is
CC   increased, the development of atherosclerosis or osteoporosis is
CC   inhibited and when the level is decreased the development of breast
CC   cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC   useful for detecting the presence of oestrogen signaling in a mammalian
CC   cell, by detecting the expression or activity of MRP or the level of MRP
CC   mRNA in the cell. The presence of oestrogen signaling is used to
CC   determine the responsiveness of the cell to the oestrogen or the tissue
CC   specific distribution of oestrogen signaling in a mammal. MRP is useful
CC   as a marker gene for the investigation of tissue-specific and oestrogen
CC   receptor-specific ligands such as oestrogen receptor antagonists and
CC   agonists. Modulators, recombinant forms, derivatives, variants or
CC   fragments of MRP are useful for enhancing or inhibiting oestrogen
CC   signaling in cells and for treatment of breast cancer or other cancers
CC   and for preventing or treating wide variety of oestrogen-related
CC   diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC   disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC   preventing and/or treating hearing loss and deafness. MRP polypeptides
CC   are useful for generating monoclonal and polyclonal antibodies useful for
CC   identifying oestrogen responsive cells, particularly cells responsive to
CC   oestrogen through an oestrogen receptor alpha or beta, to identify
CC   alleles, interspecies homologues and polymorphic variants and in
CC   immunoassays to detect MRP. The present sequence represents the coding
CC   sequence of human MRP variant #2 as described in the invention.
XX
SQ   Sequence 3780 BP, 838 A; 1229 C; 1167 G; 607 T, 0 other;
```

Query Match 21.6%; Score 946.4; DB 22; Length 3780;
Best Local Similarity 72.6%; Pred. No. 2.7e-206;
Matches 1322; Conservative 0; Mismatches 416; Indels 76; Gaps 5;

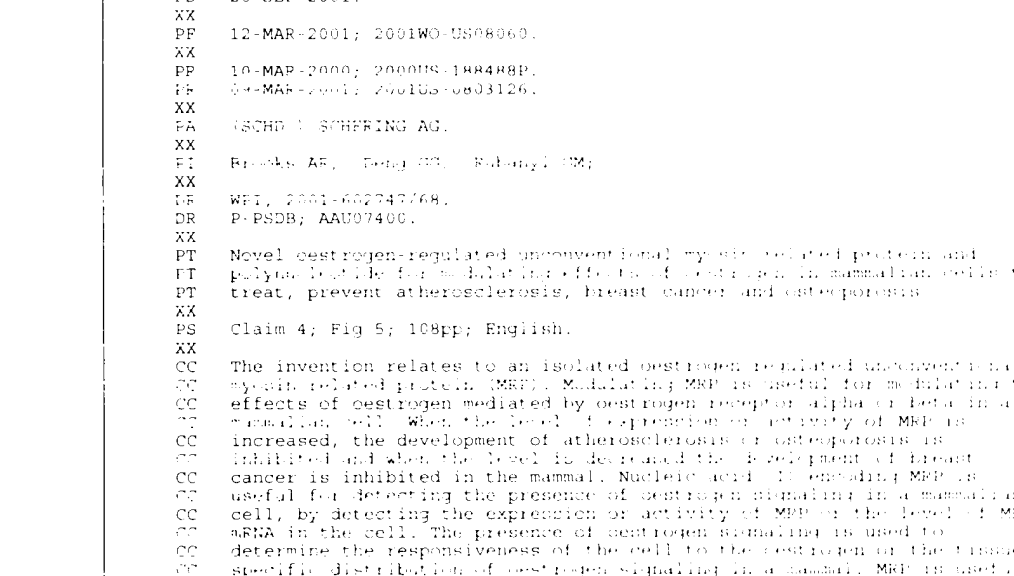


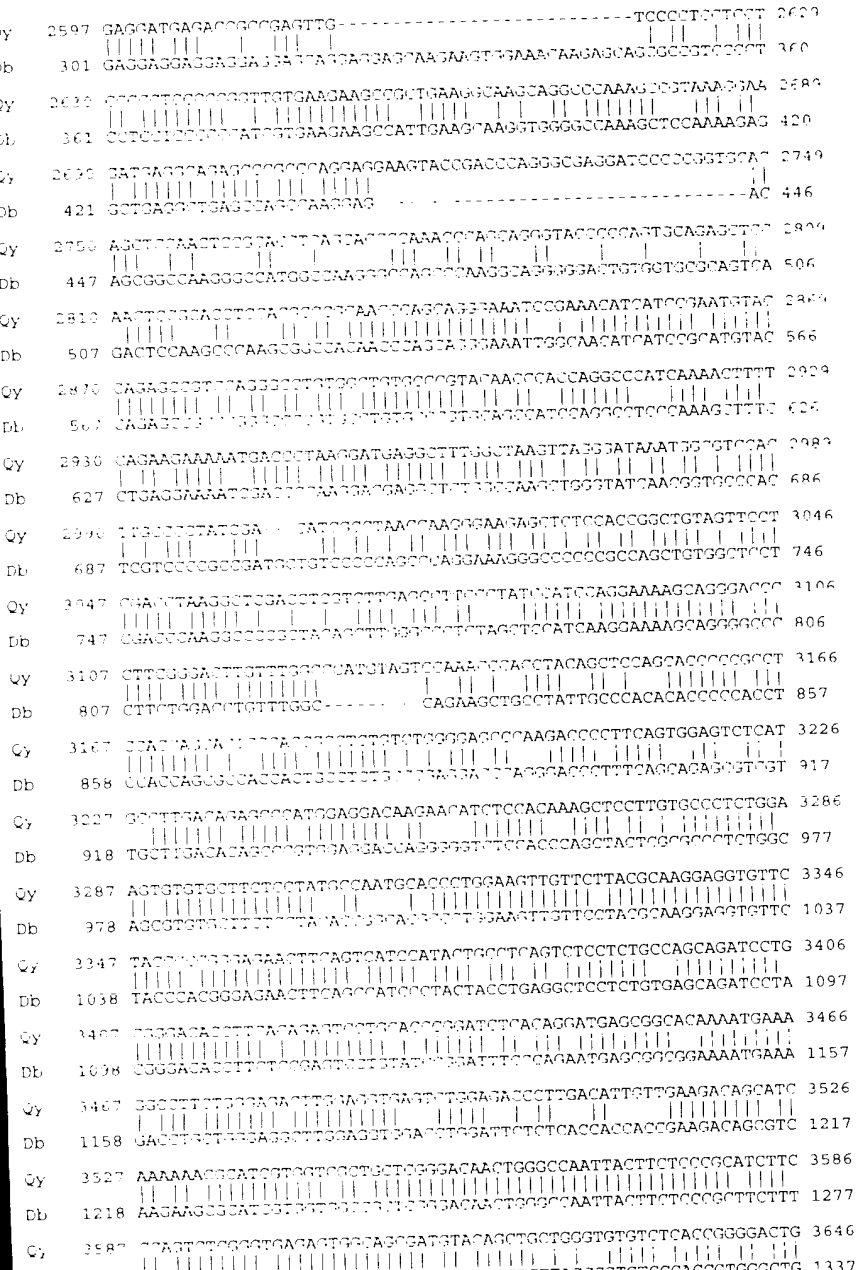


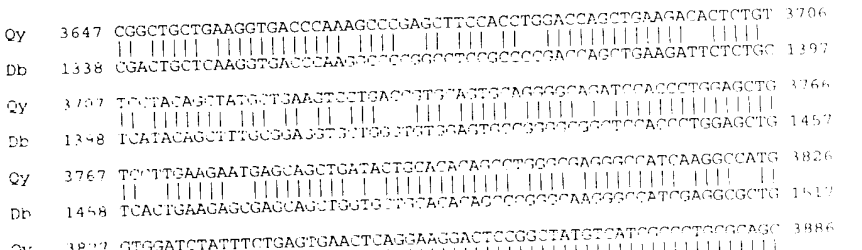


RESULT 4
AAS13821
ID AAS13821 standard; cDNA; 4174 BP.
XX
AC AAS13821;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding human myosin related protein, hMRP #1.
XX
KW Human; oestrogen regulated unconventional myosin-related protein; MRP;
KW auditory; anti-atherosclerotic; osteopathic; neuroprotective, neurotropic;
KW antiparkinsonian, cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogen;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 561..1874
FT /*tag= a
FT /product= "hMRP"
FT /note= "Oestrogen regulated unconventional myosin-
FT related protein"
XX
PN WO200168866 A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PP 10-MAR-2000; 2000US-188488P.
PR 03-MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Brooks AE, Tang CK, Rahmani M;
XX
DR WPI; 2001-602747/68.
DR P-PSDB; AAU07400.
XX
PT Novel oestrogen-regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells -
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 4; Fig 5; 108pp; English.
XX
CC The invention relates to an isolated oestrogen regulated unconventional
CC myosin related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signalling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signalling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the cDNA
CC sequence of human MRP variant #1 as described in the invention.
XX
SQ Sequence 4174 BP; 901 A; 1339 C; 1237 G; 697 T; 0 other;

Query Match 21.6%; Score 946.4; DB 22; Length 4174;
Best Local Similarity 72.9%; Pred. No. 2.8e-206;
Matches 1322; Conservative 0; Mismatches 416; Indels 76; Gaps 5;







RESULT 5
AAZ56716
ID AAZ56716 standard; cDNA; 1476 BP.
XX
AC AAZ56716;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN-19 encoding cDNA.
XX
KW Human; transmembrane protein, HTMPN, diagnosis, immunospecific;
KW antiproliferative, neuroprotective, immune disorder;
KW reproductive disorder; smooth muscle disorder, neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder, ss.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11904.
XX
PR 29-MAY-1998; 98US-0087260.
PR 02-JUL-1998; 98US-0091674.
PR 02-OCT-1998; 98US-0102954.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Bandman O, Patterson C, Gorgone GA, Fassi MF, Baughn MR;
PI Au-Young J;
XX
DR WPI; 2000-072605/06.
DR P-PSDB; AAY57895.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,

```
PT   neurological, gastrointestinal, developmental and cell proliferative
PT   disorders
XX
PS   Claim 9; Page 189; 229pp; English.
XX
CC   AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC   transmembrane proteins designated HTMPN 1 to HTMPN 79, respectively.
CC   The transmembrane protein have immunospecific, antiproliferative and
CC   neuroprotective activities. The human transmembrane proteins,
CC   polynucleotides encoding them and other compositions and methods from
CC   the present invention, can be used for the diagnosis, treatment or
CC   prevention of immune, reproductive, smooth muscle, neural glial,
CC   gastrointestinal, developmental and cell proliferative disorders. The
CC   HTMPN's can be used to treat or prevent disorders associated with a
CC   decreased expression or activity of HTMPN.
XX
SQ   Sequence 1476 BP; 277 A; 479 C; 435 G; 285 T; 0 other;

  Query Match             16.2%;  Score 706.6;  DB 21;  Length 1476;
  Best Local Similarity   78.3%;  Pred. No. 1.4e-151;
  Matches  [illegible];  Conservative    0;  Mismatches  234;  Indels   13;  Gaps    3;

Qy     2948 AAGGATGAGGCTTGGCTAAGCTAGGGATAAATGGCGTCCACTTGCCCCTATCGA---A 3004
Db        2 AAGGACGAGGCTCTGGCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTG 61

Qy     3005 T[illegible]TAACCAAGGGAAGAGCTCTCCACCGGCTGTAGTTCCTCGACCTAAGGCTCGACCT 3064
Db       62 T[illegible]CCAGCCCAGGAAAGG[illegible]CAGCTGTGGCTCCTCGACCCAAGG[illegible]CTA 121

Qy     3065 [illegible]TTTGAGCC[illegible]CCTATCCATCCAGGAAAAGCAGGGACCCCTTCGGGACTTGTTTGGC 3124
Db      122 [illegible]TCTAGCTCCATCAAGGAAAAGCAGGGGCCCCTTCTGGACCTGTTTGGC 181

Qy     3125 [illegible]CAAA[illegible]CACTTACAGCTCCAGCACCCCGGCTCCACCAGCACTCCCACCG 3184
Db      182 -CAGAA[illegible]TATTGCCCACACACCCCCACCTCCACCAGCCCCACCACTG 232

Qy     3185 [illegible]AAGA[illegible]AGTGGAGTCTCATGCTTTGACAGAGCCCATG 3244
Db      233 [illegible]GACGACCCAGGGACCCTTTCAGCACAGGTCGTTGTTTGACACAGCCCGTG 292

Qy     3245 GAGGACAAGAACATCTCCACAAAGCTCCTTGTGG[illegible]CTGGAAGTGTGTGCTTCTCCTAT 3304
Db      293 [illegible]TACCTCTGGCAGCGTGTGCTTCTCCTAC 352

Qy     3305 [illegible]AATG[illegible]CCTGGAAGTTGTTCCTTACCCAAGGAGCTGTTCTACCCCCGGGAGAACTTC 3364
Db      353 A[illegible]CCTGGAAGTTGTCCCTACGCAAGGAGGTGTTCTACCCACGGGAGAACTTC 412

Qy     3365 A[illegible]CATACTGCCTCAGTCTC[illegible]AG[illegible]CTTCACAGAG 3424
Db      413 AG[illegible]CTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAG 472

Qy     3425 T[illegible]GGAT[illegible]CACAGGATGAG[illegible]ACAAAATGAAAGG[illegible]TCTGGGAGACTTG 3484
Db      473 T[illegible]GGAT[illegible]AGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGGCTTG 532

Qy     3485 GA[illegible]GTGAG[illegible]TGGAGACCCTTGACATTGTTGAAGACAGCATCAAAAAACGCATCGTGGTC 3544
Db      533 GAGGTGGACCTGGAT[illegible]CACCACCGAAGACAGCGTCAAGAAGCGCATCGTGGTG 592

Qy     3545 GCTGCT[illegible]GGACAACTGGGCCAATTACTTCTC[illegible]GCATCTTCCCAGTCTCGGGTGAGAGT 3604
Db      593 GCTGCTCGG[illegible]ACAACTGGGCCAATTACTTCT[illegible]CTTCTTTCCTGTCTCGGGCGAGAGT 652

Qy     3605 [illegible]CCCACCGGGACTGCGGCTGCTGAAGGTGACC 3664
Db      653 [illegible]GTGCACCTGTTACCCCTGTCC[illegible]CTGGACTGCTCAAGGTGACC 712

Qy     3665 CAAA[illegible]CACCTTCCACCTGGACCAGCTGAAGACACTCTGTTCCTACAGCTATGCTGAA 3724
Db      713 CAA[illegible]GCCTCCGCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAG 772

Qy     3725 GTCCTGACCGTGCAGTGCAGGGGCAGATCCA[illegible]AAGAA[illegible] 3784
Db      773 GTCCTGGGTGTGGACTGCCGGGC[illegible]TCCACCTCGGA[illegible]AAGA[illegible] 832

Qy     3785 TTGATACTGCACACAGCCTGGGCGAGGGCCATCAAGG[illegible] 3844
Db      833 CTGGTGCTGCACACAGC[illegible]CGGGCAAGGGCCATCGAGG[illegible]TATTCCTGAAT [illegible]

Qy     3845 GAA[illegible]AAT [illegible]
Db      893 GAGCTTAAGAAG[illegible] [illegible]

Qy     3905 AG[illegible]TCAG[illegible]TTCCACC[illegible]GGGACCTCATTAGG[illegible] [illegible]
Db      953 AGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAA[illegible]A[illegible]TGGA [illegible]

Qy     3964 ACCAGGCTGGCAGTTCGGTTCTGCCGGGGGCCGCTCCGGA[illegible]GATGACGTGGT 40[illegible]
Db     1013 GCCAGGCTGGCAGTTTGGCTCTGC[illegible]GGGC[illegible]GTTCCGGA[illegible]ATA[illegible] [illegible]

Qy     4024 GCAGCCAGCTGCTGCCCCGGACCTCTCCTTTTCCCTGGGAAAGAGAAACAGCT[illegible]A 40[illegible]
Db     1073 GCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAA[illegible]A[illegible]GGCA 11[illegible]

RESULT 6
AAD07791
ID   AAD07791 standard; cDNA; 1044 BP.
XX
AC   AAD07791;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein encoding gene 8 cDNA clone HRGNC06, SEQ ID NO:[illegible]
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality, developmental abnormality, haematopoietic disorder;
KW   immune system disorder, AIDS, autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis, sepsis, diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder, kidney disorder;
KW   gastrointestinal disorder; pregnancy related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification, ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             616..657
FT                   /*tag= a
FT                   /product= "Human secreted protein precursor"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200132687-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US29364.
XX
PR   29-OCT-1999;   99US-0162238.
PR   [illegible]-JUN-2000; 2000US-0215134.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Duan RD,  Fiscella M,  [illegible]
XX
DR   WPI; 2001-308740/32.
DR   P-PSDB; AAE03399.
XX
```

```
DT   Isolated nucleic acid molecule encoding a human secreted protein is
DT   used in preventing, treating or ameliorating a medical condition -
XX
DS   Claim 1; Page 362; 398pp; English.
XX
CC   AAE03378-AAE03402 represent cDNAs corresponding to 10 human secreted
CC   protein genes, and AAE03378-AAE03402 represent the proteins they encode
CC   AAE03403-AAE03415 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   10 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease, schizophrenia, asthma,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein encoding cDNA of the invention.
XX
SQ   Sequence 1044 BP; 217 A; 350 C; 326 G; 148 T; 3 other;

  Query Match             14.4%;  Score 628.8,  DB 22;  Length 1044;
  Best Local Similarity   76.4%;  Pred. No. 7.5e-134;
  Matches  797;  Conservative    1;  Mismatches  236;  Indels    9;  Gaps    2;

Qy    1635 CGGGAGAGCACTTTGCAGGGTGGATCCTACAGAGCAGAGGCCTGGAGGCGCCCCCTCGTG 1634
           ||||    || | |  | |||| ||| |||||||||||||||||||||||||| || || |
Db       7 CGGGGAGCANTTTGGNTGAATGGACCTGCAGAGCAGAGGCCTGGAGGCGCCTCCCCGGG 66

Qy    1635 GCTGGTCTGTGTGACTGCATTCTGGGGATGCTTGGCGTGACTTGCCTGGCTGTGACTTTG 1694
           |||||||||||||||| | ||  |||| || |||  |||||| ||||||| |||||||||
Db      67 GCTGGTCTGTGTGACTCCACTCCAGGGACGCATGGCAGGACTTGGCTGGCTGCGACTTTG 126

Qy    1695 TGTTGGACCTAAATAGGCCAGACTGAGGACTTGGGAGACCCAGCTGGTCCCCACAACTACC 1754
           || |||||||| ||  |||||||||||||| |||| ||||||||| | |||| || ||||||
Db     127 TGCTGGACCTGATCAGCCAGACTGAGGACTTGGGGACCCAGCTGGCCCCCGCAGCTACC 186

Qy    1755 CCATCACTTCCTCTTTGCTTTAGCTGAGAGCATGCTTCCAGCCCCTGGTGTCCAGGCTCCTT 1814
           |||||||||||||||| | ||| |||  |||  || |  |||||||  | |||| || |||
Db     187 CCATCACTTCCTCTTTGCTTAGCCGAGGGCATTCCCCTTGCCCCTGGCATTCAGGCCCCCT 246

Qy    1815 CCCTCCCCCAGGACTCCCTTCAGGTCCAGCCCCAATACTGGCCAGCAGCCGCCCTCCGG 1874
           | |||||||||||||||||||||||||||||||||||  ||| ||||||   |   | |
Db     247 CACTGCCCCCAGGACTCCCTCCAGGTCCAGCCCCAAGCTGCCCAGCAGGGACCACACAG 306

Qy    1875 GCGAGGCCAGTAAGCCTGAGAACCTGGATGGTTTCGTGGACCACCTCTTTGAACCAGCGC 1934
           | |||| |   | | | ||| |||||| || ||| |||||||  ||||  | |||| |
Db     307 GGGAGGTCCAGAGGTCAGGGAGTTGGAGGGTTCCTGGACCAGATCTTCCAGCCAGTGA 366

Qy    1935 TCGCTCCGGGTTTCAGTGATCTGGAACAAGGTTGGGCCCTGAGCACACGCATGAAGGGAG 1994
           |   | | ||  |||| |||||||||||||||||||||||||||||||| ||||||||||||
Db     367 TATCCTCCGGCCTCAGCGATCTGGAACAAAGTTGTGTCTGAGCACGCGCATGAAGGGAG 426

Qy    1995 CCCCTCTTGTTGGGCCCACCCAGCAGGGCTACCCCATGGTGTACCCAGGTATGGTGCAGG 2054
           |||||  | |||||||||||||||||||||||||||||||||||||||||| || | |||
Db     427 GGGGCGCCATTGGGCCCACACAGCAGGGCTACCCCATGGTGTACCCAGGAATGATTCAGA 486

Qy    2055 CACCTAGCTACCAGCCAGCTATGATACCCGCACCGATGCCCGTCATGCCAGCCATGGGCG 2114
           |||   |||||||||  ||| | || |||| |||||| | |||||||| ||||||
Db     487 TGCCTGCATACCAGCCAGGCATGGTGCCTGCACCCATGCCCATGATGCCAGCAATGGGCA 546

Qy    2115 CAGTCCCAACCATGCCAGCCATGATGGTGCCACCCAGCCACACCCTCTGGTGCCCAGTT 2174
           |||||||  |||||||||||||||| |||||||  || |||||  |||||
Db     547 CAGTCCCTGCCATGCCAGCCATGGTGGTGCCGMCGCAGCCACCGC--  TTCCCAGCC 600

Qy    2175 TGGACTCAAGGCAGCTGGCACTACAGCAGCAAAACTTCATCAACCAGCAGGCGATGATTC 2234
           |||| || |||||||||  | |||||||| |||| ||| | |||||||||  | | |
Db     601 TGGATGCAGGCAGCTGGCCGTCCAGCAACAGAACTTTATCCAGCCAGCAGGCCCTAATCC 660

Qy    2235 TGGCGCAGCAGATGACCACCCAGGCCATGACCTTCTCCCCGGAGCAGCAGAATCAGAGAC 2294
           |||| |||||||||  |||||||||||||  ||||||||||||||||||  |||  |
Db     661 TGGCCCAGCAGATGACAGCCAGGCCATGTCCTTCTCCCCGGAGCAGCAGACGCAGCAGC 720

Qy    2295 ACCAGCACCAAGCTCAGACCTCTGGGGCCACCCTCCCAGCCTCCACCCTCAGCCCTCAGCCT 780
           |||| || |||| | |||| | |||  | |||||| |  |||||||  ||
Db     721 GGCAGCAGCAGGCTCGGGCCTCCGACCTTCGTCCAGGTTCCACCCTCAGCCCTCAGCCT 780

Qy    2355 CCAAGGCCAAGAAGCCTCCTGGCCCCAAGAGAAGCCAGAGAGTAACCTAAGAGCCTTCGG 2414
           |||| || |||||| ||  |||| ||||||||| || |||| |  |  ||
Db     781 CCAAGCCCAGGAAGCCCCCACACCCCCGAGAAGCCACAGCCGACCTCGGATCAGAGG 840

Qy    2415 GTGTTGGCTTGAGAGAGACAGCCAGAGAAGTGAAAGTGAACCCCAGAGCCCAAGA 2474
           |||  || ||||  || ||| | ||||| ||||  || ||| | |  |||||||
Db     841 GTGGCTGCCTGAG--AGGAGACCTCCGAGAGAGCTGAAGACAGCGCCTATCAGCCCAAGA 897

Qy    2475 GCTTCCAACAGAAACGGGACTATTTCCAGAAGATGGGCCAACATCCGATCAGAGTGAAGA 2534
           |||||| |||||||| |||||||||||  |||||||  |  |||| || |||||| ||
Db     898 GCTTCCAGCAGAAACGGAACTATTTCCAAAGGAGGGCAGCCACAGATCACAGTGAGGA 957

Qy    2535 CGGTGAAACCTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGG 2594
           || |||  || ||||||||| || |||||||| | |||   | |||||  ||||||||
Db     958 CGATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGAAGCCCAGAAGAAGGAGGAGG 1017

Qy    2595 AGGAGGATGAGACCCCCGAGTTG 2617
           ||||||| |||   |  |||  |
Db    1018 AGGAGGAGGAGGAGGAGGAGCNG 1040


RESULT 7
ABK43544
ID   ABK43544 standard; cDNA; 1268 BP.
XX
AC   ABK43544;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding novel central nervous system protein #124
XX
KW   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW   acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes, cancer, leukaemia, neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200158218-A2
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01232
XX
PR   31-JAN-2000; 2000US-0179065
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
```

PR [illegible] 2000; 2000US-0186359.
PR 10-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0199123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214980.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216889.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231965.
PR 08-SEP-2000; 2000US-0232041.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0236967.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251040.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-581633/65.
DR P-PSDB, AAU87214.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 1; SEQ ID No 134; 837pp; English.

X
C The invention describes an isolated nucleic acid molecule (I) encoding a
C novel central nervous system protein (I) and polypeptides (III) encoded
C by (I), are used to treat a medical conditions and in diagnosis of a
C pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 14.3%; Score 626.4; DB 23; Length 1268;
Best Local Similarity 80.6%; Pred. No. 2.9e-133;
Matches 732; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 3172 GCACTCCCACCGCCTGTGTCTGGGAGCCAAGACCCTTCAGTGGAGTCTCATGCCTTG 3232
Db 5 GCGTCCGCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGCTTG 64

Qy 3233 ACAGAGC [illegible] 3292
Db 65 ACACAGCCCTCCGAGACACAGAACTCTCCACAAAGTCCCTTGTGCCCTCTGGAAGTGTG 124

Qy 3293 TACTTCCCCATGCCAATGTACCCTGGAAGTTGTTCTTACGCAAGGAGGTGTTCTACCCC 3352
Db 125 TGCTTCTCCTACACCGGCACCGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCA 184

Qy 3353 CGGGAGAACTTCAGTGATCCATACTCCCTCATCTCTCCTCCCAGCAGATCCTGCGGGAC 3412
Db 185 CGGGAGAACTTCAGCCATCCTTACTACTGAGCTCTCTGTGAGCAGATCCTACGGGAC 244

Qy 3413 ACCTTCACAGAGTCCTCCACCCGGATCTCACAGGATGAGCGGCACAAAATGAAAGGCTT 3472
Db 245 ACCTTCTCCGAGTCCTGTATCCGGATTT [illegible] 304

Qy 3473 CTGGGAGACTTGGAGGTGAGTCTGGAGA [illegible] 3532
Db 305 CTGGGACGCTTTGGAGCTTACTCGAATT [illegible] 364

Qy 3533 CGCATCGTGGTCGTTGTCCTGGACAACTGGGCCAATTACTTCTCCCGCATCTTCCCAGTC 3592
Db 365 CGTATCGTGGTCGCCCCTGGACAACTGGGCCAATTACTTCTCCCGCTTCTTTCCTGTC 424

Qy 3593 TCGGGTGAGAGTGGCAGCATGTACAGCTGCTGGGTGTGTCTCACCGGGGACTGCGGCTG 3652
Db 425 TCGGGCGAGAGTGGCAGCGACGTCAGCTGTTACCCGTGTCCCACCGTGGGCTGCGACTG 484

Qy 3653 CTGAAGGTGACCCAAAGGCCTGAGCTTCCAGCTGGACCAGCTGAAGACACTCTGTTCCTAC 3712
Db 485 CTGAAGGTGA [illegible] 544

Qy 3713 AGCTATGCTGAAGTCCTGACCGTGCAGTGCA [illegible] 3772
Db 545 A [illegible] 604

Qy 3773 AAGAATGAGCAGCTGATACTGTACTACAGCTGGGCGAGCGCCATCAAGGCCATGGTGGAT 3832
Db 605 AAGAGCGAGCAGCTGGTGCTGCACACAG [illegible] 664

Qy 3833 CTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTCATCCCCCTGCGCAGCTACATC 3892
Db 665 CTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATGTCATCCCCCTGCGCAGCTACATC 724

Qy 3893 ACCGATGACAATAGCCTCCTCAGTTTCCACCGTGGGGACCTCATTAGGTTACTGCCAGTG 3952
Db 725 ACTGACAACTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCACTAAGCTACTGCCGGTG 784

Qy 3953 ACCGCTCTGGAACCAGGCTGGCAGTTCGGGT [illegible] 4012
Db 785 GCCACCCTGGAGCCAGGCTGGCAGTTTGGCTCTG [illegible] 844

Qy 4013 GATGACGTGGTGCAGCCAGCTGCTGCCCCCGACCTCTCTTTTCCCTGGGAAAGAGAAAC 4072
Db 845 GCCGACATAGTGCAGCCGGCTGCCCCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGT 904

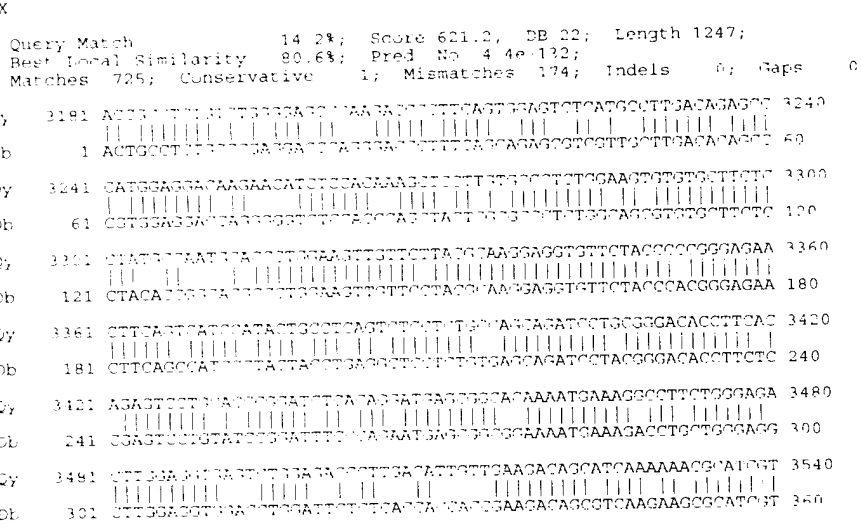
Qy 4073 AGCTGGCA 4080
Db 905 GGCTGGCA 912

RESULT 8
AAS27110
ID AAS27110 standard; cDNA, 1247 BP.
XX
AC AAS27110;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 145.
XX
KW Neuroprotective; cytostatic, dermatological, immunosuppressive, tumour;
KW antiinflammatory, anti-HIV; antibacterial, antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis, inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Alzheimer's disease;
KW neurodegenerative disorder, Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome, ischaemia, renal disorder;
KW cardiovascular; respiratory, wound healing; endocrine; Addison's disease;
KW reproductive system, gastrointestinal, liver disorder; AIDS, ss;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001, 2001WO-US01312
XX
PR 31-JAN-2000, 2000US-0179065.
PR 04-FEB-2000, 2000US-0180628.
PR 24-FEB-2000, 2000US-0184664
PR 02-MAR-2000; 2000US-0186350.
PR 15-MAR-2000, 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123
PR 19-MAY-2000; 2000US-0205515.
PR 01-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000, 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487
PR 11-JUL-2000, 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290
PR 26-JUL-2000, 2000US-0220963.
PR 26-JUL 2000; 2000US-0220964
PR 14-AUG-2000, 2000US-0224518
PR 14-AUG 2000; 2000US-0224519
PR 14-AUG-2000, 2000US-0225213.
PR 14-AUG-2000, 2000US-0225214.
PR 14-AUG-2000, 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 14-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226968.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227309.
PR 29-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 07-SEP-2000; 2000US-0229213.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230439.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232083.
PR 08-SEP-2000; 2000US-0232081.
PR 13-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234273.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 17-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250891.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-466460/50.
DR P-PSDB; AAU17193.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the prostate, bladder cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 145; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood related disorders (sickle cell anaemia, myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease, chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce

C tumour proliferation in pathologies e.g. acquired immune deficiency
C syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
C pathway protein coding sequences and PCR primers of the invention.
X

Query Match 14.2%; Score 621.2; DB 22; Length 1247;
Best Local Similarity 80.6%; Pred. No. 4.4e-132;
Matches 725; Conservative 1; Mismatches 174; Indels 0; Gaps 0;



RESULT 9
ABK43853
ID ABK43853 standard; cDNA; 1247 BP.

XX
AC ABK43853;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #433.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX
OS Homo sapiens
XX
PN WO200155318-A2
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01330
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229087.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.

PR 08 SEP 2000; 2000US-0231413.
PR 08 SEP 2000; 2000US-0231414.
PR 08 SEP-2000; 2000US-0232080.
PR 08 SEP 2000; 2000US-0232081.
PR 12 SEP 2000; 2000US-0231969.
PR 14 SEP 2000; 2000US-0232397.
PR 14 SEP-2000; 2000US-0232398.
PR 14 SEP 2000; 2000US-0232399.
PR 14 SEP 2000; 2000US-0232400.
PR 14 SEP-2000; 2000US-0232401.
PR 14 SEP-2000; 2000US-0233063.
PR 14 SEP-2000; 2000US-0233064.
PR 14 SEP-2000; 2000US-0233065.
PR 21 SEP-2000; 2000US-0234223.
PR 21 SEP-2000; 2000US-0234274.
PR 25 SEP-2000; 2000US-0234997.
PR 25 SEP 2000; 2000US-0234998.
PR 26 SEP 2000; 2000US-0235484.
PR 27 SEP-2000; 2000US-0235834.
PR 27 SEP 2000; 2000US 0235836.
PR 29 SEP-2000; 2000US-0236327.
PR 29 SEP-2000; 2000US 0236367.
PR 29 SEP 2000; 2000US-0236368.
PR 29 SEP 2000; 2000US-0236369.
PR 29 SEP 2000; 2000US-0236370.
PR 02 OCT-2000; 2000US-0236802.
PR 02 OCT-2000; 2000US-0237037.
PR 02 OCT-2000; 2000US-0237038.
PR 02 OCT-2000; 2000US-0237039.
PR 02 OCT-2000; 2000US-0237040.
PR 13 OCT-2000; 2000US 0239935.
PR 13 OCT 2000; 2000US-0239937.
PR 20 OCT-2000; 2000US-0240960.
PR 20 OCT 2000; 2000US 0241221.
PR 20 OCT-2000; 2000US 0241785.
PR 20 OCT 2000; 2000US-0241786.
PR 20 OCT-2000; 2000US 0241787.
PR 20 OCT-2000; 2000US-0241808.
PR 20 OCT 2000; 2000US-0241809.
PR 20 OCT 2000; 2000US 0241826.
PR [illegible] 2000; [illegible] 0244617.
PR 07 NOV 2000; 2000US 0246474.
PR 08 NOV 2000; 2000US-0246475.
PR 08 NOV 2000; 2000US 0246476.
PR 08 NOV 2000; 2000US 0246477.
PR 08 NOV 2000; 2000US 0246478.
PR 08 NOV 2000; 2000US 0246523.
PR 08 NOV 2000; 2000US 0246524.
PR 08 NOV 2000; 2000US 0246525.
PR 08 NOV 2000; 2000US-0246526.
PR 08 NOV 2000; 2000US-0246527.
PR 08 NOV 2000; 2000US 0246528.
PR 08 NOV 2000; 2000US-0246532.
PR 08 NOV 2000; 2000US-0246609.
PR 08 NOV 2000; 2000US 0246610.
PR 08 NOV 2000; 2000US-0246611.
PR 08 NOV 2000; 2000US-0246613.
PR 17 NOV 2000; 2000US 0249207.
PR 17 NOV 2000; 2000US-0249208.
PR 17 NOV 2000; 2000US-0249209.
PR 17 NOV 2000; 2000US 0249210.
PR 17 NOV 2000; 2000US 0249211.
PR 17 NOV 2000; 2000US 0249212.
PR 17 NOV 2000; 2000US 0249213.
PR 17 NOV 2000; 2000US-0249214.
PR 17 NOV 2000; 2000US 0249215.
PR 17 NOV 2000; 2000US 0249216.
PR 17-NOV 2000; 2000US-0249217.
PR 17 NOV 2000; 2000US 0249219.
PR 17 NOV 2000; 2000US-0249244.
PR 17 NOV 2000; 2000US 0249245.
PR 17 NOV 2000; 2000US-0249264.
PR 17 NOV 2000; 2000US 0249265.

PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17 NOV 2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR P-PSDB; AAU87523.
XX
PT New isolated nucleic acid encoding a protein for diag. use,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives
XX
PS Claim 1; SEQ ID No 443; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or lung, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malformations,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin ageing due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in bandages. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

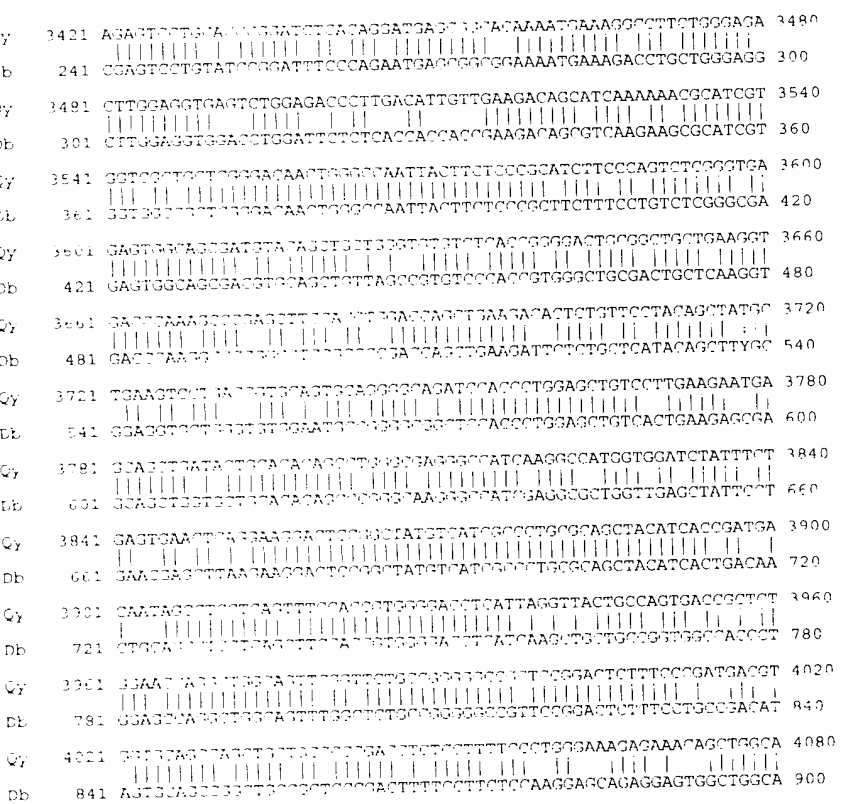
Query Match 14.2%; Score 621.2; DB 2; Length 1.47;
Best Local Similarity 80.6%; Pred. No. 4.4e-142;
Matches 725; Conservative 1; Mismatches 174; Indels [illegible]; Gaps [illegible];

Qy 3181 ACCGCCTCTGTCTGGGGAGCCCAAGACCCTTTCAGTGGAGTCTCTATCTTTGAACAAGTC 3424
Db 1 ACTGCCTCTGCCCGAGGACCCAGGCACCCTTTCAGCAGAGCGTCTCTTGACAACAGTC 60

Qy 3241 CATGGAGGACAAGAACATCTCCACAAAGTCCTTGTGTCTTTTAAATCTGTTTTTC 3300
Db 61 CGTGGAGGACCAGGGGGTCTCCACCAGCACCTCGTCCTCCTCAATCTGTTTTTC 120

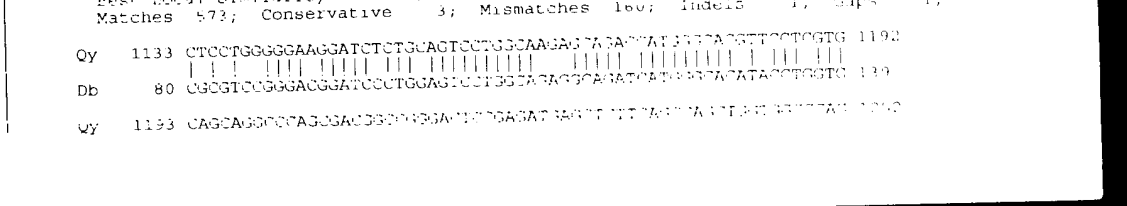
Qy 3301 CTATGCCAATGAACCTGGAACTGTCTTACCCAACGACCTCCTCATCCAGAA 3350
Db 121 CTACACGGCACCCCTGAACTTCTTACCAACGACCTCCTCATCCAGAA 170

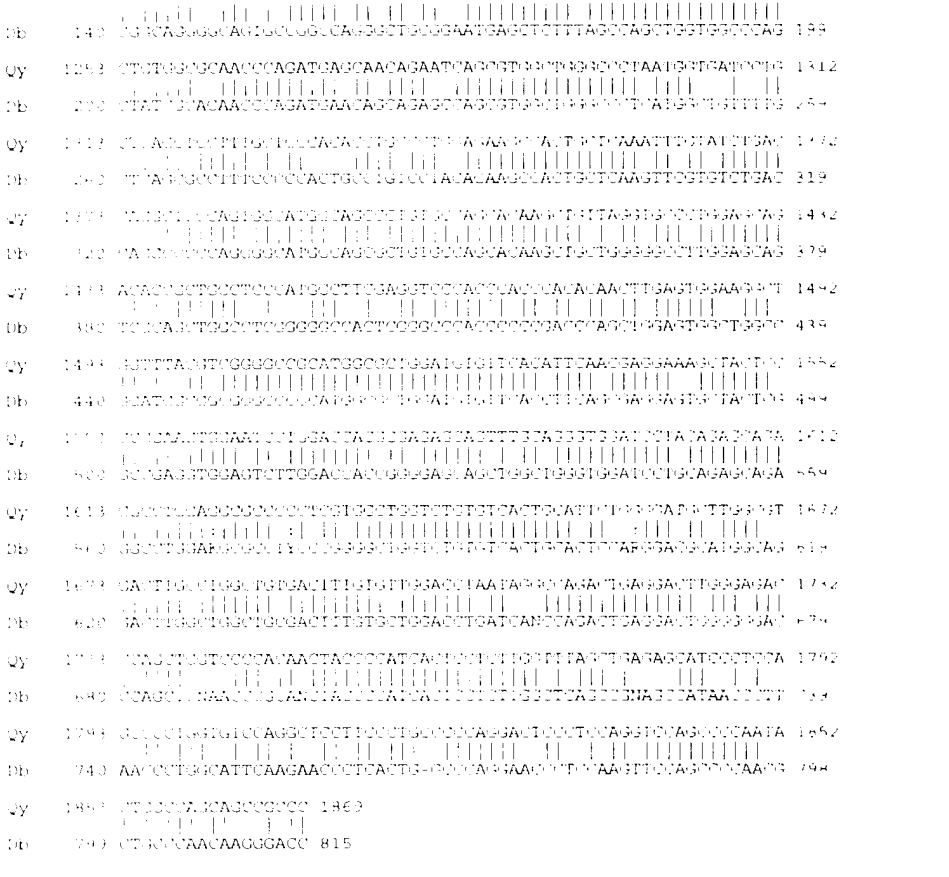
Qy 3351 CTCAGTCATCCATACTGCCTCAGTTCTTCTCCAAAAATCTCTCTA 3394
Db 181 CTTCAGCCATCCTACTACCTCAGCTCTTCTCCAAAAATCTCTCTA 224



RESULT 10
AAD07777
ID AAD07777 standard; cDNA; 826 BP
XX
AC AAD07777;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human secreted protein-encoding gene 8 cDNA clone HEONC95, SEQ ID NO:18
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder, pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerary;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   171..e93
FT   /*tag= a
FT   /product= "Human secreted protein precursor"
FT   /transl_except= (pos:567..569, aa:Xaa)
FT   /transl_except= (pos:606..608, aa:Xaa)
FT   /note= "Xaa equals any of the naturally occurring
FT   L-amino acids"
FT sig_peptide   171..293
FT   /*tag= b
FT mat_peptide   294..650
FT   /*tag= c
FT   /product= "Mature human secreted protein"
FT
XX
PN WO200132687-A1
XX
PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US29364
XX
PR 29-OCT-1999; 99US-0162238.
PR 30-JUN-2000; 2000US-0215134
XX
PA (HYMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Duan RD, Fiscella M, Ebner R;
XX
DR WPI; 2001-308740/32
DR P-PSDB; AAE03385.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition
XX
PS Claim 1; Page 351; 398pp; English.
XX
CC AAD07770-AAD07794 represent cDNAs corresponding to 10 human secreted
CC protein genes, and AAE03378-AAE03402 represent the proteins they encode.
CC AAE03403-AAE03415 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 10 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme-linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 826 BP; 150 A; 261 C; 267 G; 137 T; 11 other;

  Query Match   10.7%; Score 470.2, DB 22, Length 826;
  Best Local Similarity 77.7%; Pred. No. 1.3e-97;
  Matches 573; Conservative 3; Mismatches 166; Indels 1, Gaps 1;





RESULT 11
AAS07687
ID AAS07687 standard; DNA; 2457 BP.
XX
AC AAS07687;
XX
DT 07-NOV-2001 (first entry)
XX
DE DNA encoding novel signal transduction pathway protein, Seq ID 1947
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW [illegible] system disorder, rheumatoid arthritis, inflammatory condition,
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder, Alzheimer's disease, Parkinson's disease,
KW chromosomal abnormality, Down syndrome, ischaemia, renal disorder,
KW cardiovascular, respiratory, wound healing, addiction, Alzheimer's disease,
KW reproductive system, gastrointestinal; liver disorder; AIDS; ds;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200144771-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186490.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190276.
PR 18-APR-2000; 2000US-0198173.
PR 14-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0242080.
PR 08-SEP-2000; 2000US-0242081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0234063.
PR 14-SEP-2000; 2000US-0234064.
PR 14-SEP-2000; 2000US-0234065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802
PR 02-OCT-2000; 2000US-0237037
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039
PR 02-OCT-2000; 2000US-0237040
PR 13-OCT-2000; 2000US-0239835
PR 13-OCT-2000; 2000US-0239837.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246607.
PR 08-NOV-2000; 2000US-0246610
PR 08-NOV-2000; 2000US-0246611
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207
PR 17-NOV-2000; 2000US-0249208
PR 17-NOV-2000; 2000US-0249209
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211
PR 17-NOV-2000; 2000US-0249212
PR 17-NOV-2000; 2000US-0249213
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256713.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251890
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0256678.
XX
PA (HUMA) HUMAN GENOME SCI INC
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465463/50.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1, SEQ ID NO 1247; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX
SQ Sequence 24757 BP; 5183 A, 7437 C, 7109 G; 4669 T; 0 other;

Query Match 3.9%; Score 166.2; DB 22; Length 24757;
Best Local Similarity 80.2%; Pred No. 1.8e-27;
Matches 195; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy    1119 GGGTGATGCTGGGGCTCCTGGGGGAAGGATCTTTGCAGTTTTTGCAAAGCAGACCATGG 1178
           ||  |||||||||||||||||||| ||||| ||| |||||||||    ||||| |||||
Db   17652 GGCAGATGCTGGGGCTTCTGGGGGACGGATCTTTTGAGTTTTTGCAAAGCAGATCATGG 17593

Qy    1179 GCAGGTTTCTTGGTGCAGCAAGGCCAGGAGGGCGAGCTTTAGATAATTTTTCAGCC 1238
           || | | ||| |||| |||||  ||| | ||||| || || ||  |||||||| ||||
Db   17592 GCTGATACTCTGGTGCGGTAGAGGTAGTTGTTGCTAGGGTTGTTGAATAAGTCTTTAGCC 17533

Qy    1239 AGCTGGTGGCCCAGCTGTGGCGCAAGCCAGATGAGCAACAGAATCAGCGTGGCTGGGCCC 1298
           |||||||||||||| ||||  |||||||||||| || ||||  ||||||||||||||||
Db   17532 AGCTGGTGGCCCAGCTATGGCAGAAGCCAGATGAACAGAGATCAGCGTGGCTGGGCCC 17473

Qy    1299 TAATGGTGATGCTGCTCAGCTCCTTTGCTCCCAAGAGTGGTGGAGAAGCCACTGCTCA 1358
           | ||||   | |||||||| ||||| | ||    |||| ||| ||||||||||||||
Db   17472 TCATGGTGTTTTTGTTCAGCGCCTTTGCCCACTGCCTTGGTACAGAAGCCACTGCTCA 17413

Qy    1359 AAT 1361
           |||
Db   17412 AGT 17410
```

RESULT 12
AAS33481
ID AAS33481 standard; DNA; 24757 BP.
XX
AC AAS33481;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 764.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW antithrombotic; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;

KW thrombosis; wound healing; ds.
XX
OS Homo sapiens.
XX
PN WO200134[illegible] A2.
XX
PD 3[illegible] AUG 2001.
XX
PF 17 JAN-2001; 2001WO-US01347.
XX
PR 31 JAN 2000; 2000US-0179065.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451931/48.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions
XX
PS Disclosure; SEQ ID No 764; 763pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC [illegible] antagonists' of the expression and activity of the secreted
CC protein. The anti-(II) antibodies and antagonists may also be used to
CC down-regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanoma, neoplasms of the breast or liver, Sezary syndrome and
CC Sanchez's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot Marie Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC anaemia and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS33045 AAS33046 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 24757 BP; 4809 A; 7328 C; 7437 G; 5183 T; 0 other;

Query Match 3.8%; Score 166.2; DB 22; Length 24757;
Best Local Similarity 80.2%; Pred. No. 1.8e-27;
Matches 198; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 1219 [illegible] 1278
Db 7165 [illegible] 7224

Qy 1279 [illegible] 1338
Db 7225 [illegible] 7284

Qy 1339 [illegible] 1398
Db 7285 [illegible] 7344

Qy 1399 TAATGGTCATCCTGCTCAGCTCCTTTGCTCCCACACCTGGGGTCTCAAAAGCCACTGCTTCA 1458
Db 7296 TCATGCCTGTTTTGCTCAGCGCTTTCCCCACTGCTCGTCTTATAAAAGCACTGTTTA 7441

Qy 1359 AAT 1361
Db 7346 AGT 7348

RESULT 13
ABA60390/c
ID ABA60390 standard; DNA; 526 BP.
XX
AC ABA60390;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #8[illegible].
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ds.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US03069.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633066.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000US-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-482447/52.
XX
PT Human genome derived single exon nucleic acid probes useful for
PT analysing gene expression in human foetal liver
XX
PS Claim 1; SEQ ID NO 8[illegible]; [illegible]pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes are/are used for production,
CC measuring and displaying gene expression profiles derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 526 BP; 85 A; 192 C; 156 G; 93 T; 0 other;

Query Match 2.0%; Score 89.4; DB 22; Length 526;
Best Local Similarity 86.1%; Pred. No. 1.5e-11;
Matches 99; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

Qy 3855 [illegible] 3914
Db 415 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCAAGAAGCGCACCCTCTA 356

Qy 3915 GTTCCACCCTGGGGACCTCATTAGGTTACTGCTACGGACCCTGTCAAC 3961
Db 355 GCTTCCACCGTGGGGACCTCATCAAGCTGCTGCAGGTGGTCATCAAG 309

```
ESULT 14
ABA28621/c
D   ABA28621 standard; DNA; 526 BP.
X
C   ABA28621;
X
T   23-JAN 2002  (first entry)
X
E   Probe #7087 for gene expression analysis in human heart cell sample.
X
W   Human; gene expression; heart; microarray; vascular system; probe;
W   cardiovascular disease; hypertension; cardiac arrhythmia;
W   congenital heart disease; ss.
X
S   Homo sapiens.
X
N   WO200157274-A2.
X
D   09-AUG-2001
X
F   30-JAN 2001; 2001WO-US00666
X
R   04-FEB-2000; 2000US-0180312
R   26-MAY-2000; 2000US-0207456
R   30-JUN-2000; 2000US-0608408
R   03-AUG-2000; 2000US-0632366
R   21-SEP-2000; 2000US-0234687.
R   27-SEP-2000; 2000US-0236359
R   04-OCT-2000; 2000GB-0024263.
X
A   (MOLE-) MOLECULAR DYNAMICS INC.
X
I   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
X
R   WPI; 2001-483446/52.
X
T   Single exon nucleic acid probes for analyzing gene expression in human
T   hearts -
X
S   Claim 1; SEQ ID NO 7187; 526pp; English
X
C   The present invention relates to single exon nucleic acid probes for
C   measuring human gene expression in a sample derived from human heart. The
C   present sequence is one such probe. The probes may be used for
C   predicting, measuring and displaying gene expression in samples derived
C   from the human heart via microarrays. By measuring gene expression, the
C   probes are useful for predicting, diagnosing, grading, staging,
C   monitoring and prognosing diseases of the human heart and vascular system
C   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
C   congenital heart disease.
C   Note: The sequence data for this parent did not form part of the printed
C   specification, but was obtained in electronic format directly from WIPO
C   at ftp.wipo.int/pub/published_pct_sequences
X
Q   Sequence 526 BP; 85 A; 192 C; 166 G; 83 T; 0 other;

  Query Match                 2.0%;  Score 89.4;  DB 22;  Length 526;
  Best Local Similarity   86.1%;  Pred. No. 1.5e-10;
  Matches   99;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    3855 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACCGATGACAATAGCCTCCTCA 3914
           ||||||||||||||||||||||||||||||||||||||||||||||||||  ||   |||||||||
Db     415 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCA 356

Qy    3915 GTTTCCACCGTGGGGACCTCATTAGGTTACTGCCAGTGACCGCTCTGGAACCAGG 3969
           |||||||||||||||||||| | | | ||||| ||| || | ||||| |||||
Db     355 GCTTCCACCGTGGGGACCTCATTAAGCTTCGTCGCCAGCCTGGAGCCAGG 301


RESULT 15
AAK08667/c
ID   AAK08667 standard; DNA; 526 BP.
XX
AC   AAK08667;
XX
DT   05-NOV 2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 8658.
XX
KW   Human; brain expressed exon, gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667
XX
PR   04-FEB-2000; 2000US-0180312
PR   26-MAY-2000; 2000US-0207456
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 8658; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 526 BP; 85 A; 192 C; 166 G; 83 T; 0 other;

  Query Match                 2.0%;  Score 89.4;  DB 22;  Length 526;
  Best Local Similarity   86.1%;  Pred. No. 1.5e-10;
  Matches   99;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    3855 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACCGATGACAATAGCCTCCTCA 3914
           ||||||||||||||||||||||||||||||||||||||||||||||||||  ||   |||||||||||
Db     415 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCA 356

Qy    3915 GTTTCCACCGTGGGGACCTCATTAGGTTACTGCCAGTGACCGCTCTGGAACCAGG 3969
           | |||||||||||||||||| | | | ||||| ||| || | ||||| |||||
Db     355 GCTTCCACCGTGGGGACCTCATTAAGCTTCGTCGCCAGCCTGGAGCCAGG 301


Search completed. January  9, 2003, 07:15:15
Job time : 921 196 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: January 9, 2003, [illegible] ; Search Time 5202.42 Seconds
(without alignments)
13567.518 Million cell updates/sec

Title: US-09-803-126-3
Perfect score: 4375
Sequence: 1 cgctgggactgtcacctatc[illegible] aaaaaaaaaaaaaaaaaaaa 4375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	838	19.2	875	14	BQ941051	BQ941051 AGENCOURT
2	774.2	17.7	925	14	BQ960468	BQ960468 AGENCOURT
3	682	15.6	947	14	BQ941605	BQ941605 AGENCOURT
4	668.8	15.3	932	12	BF584026	BF584026 602096267
5	657.6	15.0	677	10	BB072116	BB072116 BB072116
C 6	574	13.1	610	13	BI851062	BI851062 imageqc_1
7	556	12.7	958	12	BF582477	BF582477 602094438
8	552.2	12.6	653	10	BB624737	BB624737 BB624737
9	536	12.3	929	12	BF585148	BF585148 602101892
10	504.4	11.5	1088	14	BM906425	BM906425 AGENCOURT
11	487.6	11.1	720	13	BG964584	BG964584 [illegible]
12	484	11.1	505	9	AI891960	AI891960 ui57a11.y
C 13	477	10.9	698	13	BI294877	BI294877 UI-R-DK0
14	430.2	9.8	603	14	BM791149	BM791149 K-EST00071
15	421	9.6	422	14	BQ938810	BQ938810 AGENCOURT
16	420	9.6	431	10	BB738897	BB738897 BB738897
17	416.8	9.5	642	10	BB666962	BB666962 BB666962
18	413.6	9.5	969	14	BM804816	BM804816 AGENCOURT
19	412.2	9.4	851	12	BF583262	BF583262 602101725
20	405	9.3	406	10	BB738401	BB738401 BB738401
C 21	403.8	9.2	581	10	BE116910	BE116910 UI-R-BS1-
22	400.8	9.2	1070	13	BI766913	BI766913 [illegible]
23	393.4	9.0	407	10	BB739631	BB739631 BB739631
24	382.6	8.7	389	9	AI956557	AI956557 wl28g07.y
25	370.2	8.5	445	10	BB847500	BB847500 BB847500
26	351.4	8.0	366	14	BQ126699	BQ126699 [illegible]
27	351	8.0	396	10	BB738178	BB738178 BB738178
28	348	8.0	530	13	BI007177	BI007177 QV3-RT007
29	341.4	7.8	574	14	BM738080	BM738080 K-EST0002
30	341.4	7.8	661	13	BI769659	BI769659 [illegible]
31	334.8	7.6	500	10	BE246969	BE246969 TCBAP1E60
32	329.6	7.5	1004	14	BM804831	BM804831 AGENCOURT
C 33	327.2	7.5	342	14	BQ126442	BQ126442 [illegible]
34	326.4	7.5	585	14	BM737684	BM737684 K-EST0000
35	320.8	7.3	526	13	BM193191	BM193191 TCBAP1E10
36	317.4	7.3	343	10	BB843598	BB843598 BB843598
37	298.4	6.8	319	10	BB208242	BB208242 BB208242
C 38	292.8	6.7	545	9	AI767579	AI767579 wh23d11.x
39	288.2	6.6	512	14	BM738296	BM738296 K-EST0004
C 40	286.6	6.6	442	12	BG376716	BG376716 UI-R-CU0
41	280.8	6.4	422	13	BM152526	BM152526 TCBAP1E91
42	280.8	6.4	722	13	BI772088	BI772088 604059167
43	273.4	6.2	365	10	BB843880	BB843880 BB843880
44	261	6.0	417	13	BG994301	BG994301 CM3-HT115
45	258.6	5.9	337	10	BB204202	BB204202 BB204202

ALIGNMENTS

RESULT 1
BQ941051
LOCUS BQ941051 875 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8984662 NCI_CGAP_Ov24 Mus musculus cDNA clone IMAGE:6476207 5', mRNA sequence.
ACCESSION BQ941051
VERSION BQ941051.1 GI:22356529
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 875)
AUTHORS NIH_MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14016 Row: c Column: 24
High quality sequence stop: 634.
FEATURES Location/Qualifiers
source 1..875

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6476207"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon, Vector: pCMV-SPORT6, Site_1: NotI;
Site_2: SalI, Cloned unidirectionally, Primer: Oligo: 3T
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 165 a 276 c 261 g 172 t 1 others
ORIGIN

Query Match 19.2%; Score 838; DB 14; Length 875;
Best Local Similarity 99.1%; Pred. No. 3.7e-136;
Matches 858; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

Qy 1049 CTTGACTGGTTTAGATGGGAGAAGCCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCT 1108
Db 1 CCTGACGGGTTAGATGGGAGAAGCCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCT 60

Qy 1109 GGGGCTCTGGGGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCAGCTTCCT 1168
Db 61 GCGGCTCCTGGGGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCAGCTTCCT 120

Qy 1169 CTTCAACCAGCTCTACGAGTGGATTGGAACTTCGACATGACGCTCTTCAGCTCAGCTGGTGGC 1228
Db 121 CTTCAGCAGGCTCAGCGAGTGGTGGGACTCCGAGATGAGCTCTTCAGCTCAGCTGGTGGC 180

Qy 1229 CAGCTCTGCCCCAACCCAGATGAGCAACAGAATCAGCGTGGCTGGGCCCTAATGGTGAT 1288
Db 181 CAGCTCTGCCCCAACCCAGATGAGCAACAGAATCAGCGTGGCTGGGCCCTAATGGTGAT 240

Qy 1289 CTTGCTTAGCTCCTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATC 1348
Db 241 CCTGCTTAGCTCCTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATC 300

Qy 1349 TTACCAGGCTCCCAGTGGCATGCCAGCCCTGTGCCAGCACAAGCTGTTAGGTGCCCTGGA 1408
Db 301 TTACCAGGCTCCCAGTGGCATGCCAGCCCTGTGCCAGCACAAGCTGTTAGGTGCCCTGGA 360

Qy 1409 TAGACAGCGCTGGCTCCCATGGCTTCGAGGTCCCACCCACCCACACAACTTGAGTGGAA 1468
Db 361 TAGACAGCGCTGGCTCCCATGGCTTCGAGGTCCCACCCACCCACACAACTTGAGTGGAA 420

Qy 1469 TCTTGTTAGGTGGGGCCCAAGGGCTGGATGTGTTCACATTCAAGGAGGAAAGCTA 1528
Db 421 GCTTGTTTACGTGGGGCCCCATGGGCCTGGATGTGTTCACATTCAAGGAGGAAAGCTA 480

Qy 1529 TAAGTGGAATCCTGGACCACGGGAGAGCAGTTTGCAGGGTGGATCCTACAGAG 1588
Db 481 CTCCGCGGAAGTGGAATCCTGGACCACGGGAGAGCAGTTTGCAGGGTGGATCCTACAGAG 540

Qy 1589 AAAGGCCTGGAGGCGCCCCCTCGTGGCTGGTCTGTGTCACTGCATTCTGGGGATGCTTG 1648
Db 541 CAGAGGCTGGAGGCGCCCCCTCGTGGCTGGTCTGTGTCACTGCATTCTGGGGATGCTTG 600

Qy 1649 TCTGAGTTTGCCTGGCTGTGACTTTGTGTTGGACCTAATAGGCCAGACTGAGGACTTGGG 1708
Db 601 GCGTGACTTGCCTGGCTGTGACTTTGTGTTGGACCTAATAGGCCAGACTGAGGACTTGGG 660

Qy 1709 AAACCAACTGGTCCCCACAACTACCCCATCGCTCCTCTTGGTTTAGCTGAGAGCATCCC 1768
Db 661 AGACCCAGCTGGTCCCCACAACTACCCCATCGCTCCTCTTGGTTTAGCTGAGAGCATCCC 720

Qy 1769 TCTAGTCCCTGGTGTCCAGGCTCCTTCCCTGCCCCCAGGACTCCCTCCAGGTCCAGCCCC 1828
Db 721 TCTAGTCCCTGGTGTCCAGGCTCCTTCCCTGCCCCCAGGACTCCCTCCAGGTCCAGCCCC 780

Qy 1829 AATACTGGCCAGCAGCCGCCCT-CCGGGCGAGGCCAGTAAGCCTGAGAACCTGGATGGTT 1887
Db 781 GATA.....GAAAGGGGGGGGGGAAAAAGGGAAGGGT 840

Qy 1888 TCTTGAACCACCTCTTTGAACCAGCCTCGTCCG 1941
Db 841 TCCGGGACCANCTCTTTGAAACAGCGTCGTCCG 875

RESULT 2
BQ960468
LOCUS BQ960468 925 bp "mRNA linear EST" 17-OCT-
DEFINITION AGENCOURT_8950848 NCI_CGAP_Co24 Mus musculus cDNA clone
IMAGE:6474896 5', mRNA sequence.
ACCESSION BQ960468
VERSION BQ960468.1 GI:22375946
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 925)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian gene collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cooke Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14012 row: m column: 09
5' (G) pair; sequence stop: 671.
FEATURES Location/Qualifiers
source 1..925
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6474896"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon, Vector: pCMV-SPORT6, Site_1: NotI;
Site_2: SalI, Cloned unidirectionally. Primer: Oligo: 3T
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 174 a 287 c 290 g 174 t
ORIGIN

Query Match 17.7%; Score 774.2; DB 14; Length 925;
Best Local Similarity 96.9%; Pred. No. 4.7e-125;
Matches 860; Conservative 0; Mismatches 28; Indels 9; Gaps 4;

Qy 791 GCCCCTGGCCCGTACAGCGGGAGCCCTGGGGAAAGTGTCAAATATGAACTTGTTGT 850
Db 1 GCCCCTGGCCCGTACAGCGGGAGCCCTGGGGGAAAGTGTCAAATATGAACTTGTTGT 60

Qy 851 TTAGAGATCCCGGCCAGCTGGTACTCTGCTGAAGAGCCGAAGCCAGCCTTTGTGCTT 910
Db 61 TTAGAGATCCCGGCCAGCTGGTACTCTGCTGAAGAGCCGAAGCCAGCCTTTGTGCTT 120

Qy 911 CTGACGGGGAGCATCACAGAGTCCCTGCCACCTGAGGTCCCCCCCCCAGCCTGACT 967
Db 121 CTGACGGGGAGCATCACAGAGTCCCTGCCACCTGAGGTCCCCCCCCAGCCTGACT 180

Qy 971 CTCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTGTATCACAGTTTCAGAAT 1030
Db 181 CTCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTGTATCACAGTTTCAGAAT 240

Qy 1031 CCATTCTGCCTGACCACAGGGCAGCCACCGGACCAGCCCCTGACCCCTTAATCATCCCAT 1090
Db 241 CCATTCTGCCTGACCACAGGGCAGCCACCGGACCAGCCCCTGACCCCTTAATCATCCCAT 300

Qy 1091 AATCTCTCATCAGTCTCTCATCATCTAAGTCATCTCCTCCTCCTTAAGTCATCTT 1150

```
Db     301 AACCCTCAGAGAAGTCTGGAGATTAATAGGTTAGTCTGGGGCTCCTGGGGAAGTATCT 360

Qy    1151 CTGCAGTCTTGGCAAGAGCAGACATGGCAGTTCCTCTGTGCAGCAGGCCCAGCGACCG 1210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGCAGTCCTGGCAAGAGCAGACATGGCAGTTCCTCTGTGCAGCAGGCCCAGCGACCG 420

Qy    1211 CCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAGCTGTGGCGCAACCCAGAT 1270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAGCTGTGGCGCAACCCAGAT 480

Qy    1271 GAGTAACAGAATCAGTGGTGTTGGTCCTAATGGTGATCCTGCTCAGCTCCTTTGCTCCC 1330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAGTAACAGAATCAGTGGTGTTGGTCCTAATGGTGATCCTGCTCAGCTCCTTTGCTCCC 540

Qy    1331 ACAGCTGTCCTTGAGAAGTCACTGTTCAAATTTGTATCTGACCAGGCTCCCAGTGGCATG 1390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ACAGCTGTCCTTGAGAAGTCACTGTTCAAATTTGTATCTGACCAGGCTCCCAGTGGCATG 600

Qy    1391 GCAGCCGTTTCTCAGTACAAGTTGTTAGGTGCCCTGGAGCAGACACCGCTGGCTCCCATG 1450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GCAGCCGTTTCTCAGTACAAGTTGTTAGGTGCCCTGGAGCAGACACCGCTGGCTCCCATG 660

Qy    1451 GCTTCGAGGTCCCACCCAGCCAGACAAGTTT GAGTGGAAGGCTGGTTTACGTC-GGGGC 1507
           ||||||||||||||||||||||||||||||   ||| ||||||||||| |||| |||||
Db     661 GCTTCGAGGTCCCACCCAGCCAGACAACTTTGAGTGGGAGGGCTGGTTTACGTCGGGGGC 720

Qy    1508 TTTGATGGCTTTGATGATTTCTTCAGATTCAACGAGGAAAGCTACTCCGCGGAAGTGGAATC 1566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CGCATCGTCGCTTCGATGTTTCTTCAGATTCAACGAGGAAAGCTACTCCGCGGAAGTGGAATC 780

Qy    1567 TTTGACGACTTTGATGATGACTTTGCAGGGTGGATCCTACAGAGCAGA-GGCCTGGAGGC 1623
           ||||||||||||||||||||||   ||||||||||||||||||||| ||| ||||||||||
Db     781 CTTGATGATGATGATGACTTTTTCAGGGGTGGATCCTACAGAGCAGAGGGCCTGGAGGC 840

Qy    1624 GCTGTC-TCGTCGTCGTCGTCATGCATTCTGGGATGCTTGGCGTGACTTG 1678
           ||||||  | |||||||||| |  || | | | | | | | | | | ||
Db     841 GCTGTCTTCTTCGTCGTCATTGGCATTCCTGGGGGAATGCCTTGGCGTG 897

RESULT 3
BQ941605
LOCUS       BQ941605                 947 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  ACENSORED 6475776 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6475749 5', mRNA sequence.
ACCESSION   BQ941605
VERSION     BQ941605.1  GI:32867083
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: The Cupke Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I M A G E Consortium (LLNL)
            DNA Sequencing by: Advanced Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I M A G E Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14014  row: p  column: 22
            High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6475749"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon, Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 Kb.  Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      175 a    312 c    290 g    161 t      2 others
ORIGIN

  Query Match          15.6%;  Score 682;  DB 14;  Length 947;
  Best Local Similarity   96.2%;  Pred. No. 5.3e-109;
  Matches  769;  Conservative    0;  Mismatches   30;  Indels    2;  Gaps    2;

Qy     787 GTGTGCCCCTGGCCCGCACAGCGGGGAGCCGTGGGGAAAGTGTGAAAGATGTGACCTGGG 846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32 GTGTGCCCCTGGCCCGCACAGCGGGGAGCCGTGGGGAAAGTGTGAAAGATGTGACCTGGG 91

Qy     847 TCGCTTAGAGATCCCCGCCCAGCTGGCTACTCTGCTGGAGATGTCTGAAGTCTGACAGGC 906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92 TCGCTTAGAGATCCCCGCCCAGCTGGCTACTCTGCTGGAGATGTCTGAAGTCTGACAGGC 151

Qy     907 CTTGCTGACGGGGAGCATCACAGAGTCGCTGTCACTGATGTCTGTCTGTCTAGCCT 966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152 CTTGCTGACGGGGAGCATCACAGAGTCGCTGTCACTGATGTCTGTCTGTCTAGCCT 211

Qy     967 GACTCTCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCA 1026
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 GACTCTCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCA 271

Qy    1027 GAAGCCATTTCTGCCTCGACCAGGGTACACCACTGGAAGAGGTTTGACTTTGGTAGATGG 1086
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 GAAGCCATTTCTGCCTCGACCAGGGCAGCCACTGGAAGAGGTTTGACTTTGGTAGATGG 331

Qy    1087 CGAGAACCCTCAGCAGGCTCTGGAGATCAACAGGGTGATGTCTTTTTCTTTTGGGAAGG 1146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 CGAGAACCCTCAGCAGGCTCTGGAGATCAACAGGGTGATGTCTTTTTCTTTTGGGAAGG 391

Qy    1147 ATCTCTGCAGTCCTGGCAAGAGCAGACATGGCAGTTCCTCTGTGCAGCAGGCCCAGCG 1206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 ATCTCTGCAGTCCTGGCAAGAGCAGACATGGCAGTTCCTCTGTGCAGCAGGCCCAGCG 451

Qy    1207 ACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAGCTGTGGCGCAACCC 1266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     452 ACGGCCGGGACTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAGCTGTGGCGCAACCC 511

Qy    1267 AGATGAGTAACAGAATCAGTGGTGTTGGTCCTAATGGTGATCCTGCTCAGCTCCTTTGC 1326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     512 AGATGAGTAACAGAATCAGTGGTGTTGGTCCTAATGGTGATCCTGCTCAGCTCCTTTGC 571

Qy    1327 TCCCACAGCTGTCCTTGAGAAGTCACTGTTCAAATTTGTATCTGACCAGGCTCCCAGTGG 1386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     572 TCCCACAGCTGTCCTTGAGAAGTCACTGTTCAAATTTGTATCTGACCAGGCTCCCAGTGG 631

Qy    1387 CATGGCAGCCCTGTG-CCAGCACAAGCTGTTAGGTGCCCTGGAGCAGACACCGCTGGCTC 1445
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db     632 CATGGCAGCCCTGTGCCCAGCACAAGCTGTTAGGTGCCCTGGAGCAGACACCGCTGGCTC 691

Qy    1446 CCAT-GGCTTCGAGGTCCCACCCAGCCAGACAAGTTTGAGTGGAAGGCTGGTTTACGTC 1502
           |||| |||||||||||||||| ||||||  || |  |||||||||||||||||||||||
Db     692 CCATGGGCTTCGAGGTCCCACCCCCCCACAACCACTTGAGTTGGAAGGCTGGTTTACGGC 751

Qy    1503 GGGCCGCATGG---CGCTGGATGTGTTCACATTCAACGAGGAAAGCTACTCCGC-GGA 1557
           |||| | | |   |||||||||||||||||||||||||||||||||||||||| ||| 
Db     752 GGGGGCCCCCTGGCCGCCTGGATGGGCTCCCTTCCACGAAGGAAAGCTACTCCGCCGGA 811

Qy    1558 AGTGGAATCCTGGACCACGGGAGAGCAG 1585
           |||||||||| |  | | || ||| ||
Db     812 AGTGGAATCCCTGGACCCCGGGGAGAAG 839

RESULT 4
BF584026
LOCUS       BF584026                 932 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  602096267F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4216312 5',
```

```
            mRNA sequence.
ACCESSION   BB984026
VERSION     BB984026.1  GI:11657744
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 932)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9793  row: e  column: 17
            High quality sequence stop: 690.
FEATURES             Location/Qualifiers
     source          1..932
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4216312"
                     /clone_lib="NCI_CGAP_Cu24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; Primer: Oligo dT
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP library."
BASE COUNT      240 a    244 c    284 g    174 t
ORIGIN

  Query Match             15.2%;  Score 669.8;  DB 10;  Length 932;
  Best Local Similarity   97.2%;  Pred. No. 2.1e-196;
  Matches  701;  Conservative    0;  Mismatches   17;  Indels    4;  Gaps    3;

Qy    1641 GGAGTGGTGGAGTGGTGAAGGTGAGTGAGTAAAGGAGGAAGTTGCAGTTGGACTAGCTGGAAGACA 1700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGAGTGGTGGAGTGGTGAAGGTGAGCGAGTAAAGGAGGAAGTTGCAGTTGGACTAGCTGGAAGACA 60

Qy    1701 CTTCTGTTCCTACAGCTATGCTGAAGTGCTGACCCTGCAGGGGGAGATCCAGCTG 1760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTCTGTTCCTACAGCTATGCTGAAGTGCTGACCCTGCAGGGGGAGATCCAGCTG 120

Qy    1761 GAGTGTCCTTGAAGAATGAGCAGCTGATACTGCACACAGCCTGGGCCAGGGCCATCAAG 1820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAGTGTCCTTGAAGAATGAGCAGCTGATACTGCACACAGCCTGGGCCAGGGCCATCAAG 180

Qy    1821 GCCATGCTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTGTCATGGCCTG 1880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCATGCTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTGTCATGGCCTG 240

Qy    1881 CAGGTACATCATCGGATGACAATAGCCTCCTCATTTCCACCGTGGGGACCTCATTATGG 1940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGTAGCTACATCACCGATGACAATAGCCTCCTCATTTCCACCGTGGGGACCTCATTATGG 300

Qy    1941 TTACTGTCAGTGACCGCTCTGGAACCAGGTTGGCAGTTCGGTTCTGCCGGGGGCGCTCC 2000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTACTGTCAGTGACCGCTCTGGAACCAGGTTGGCAGTTCGGTTCTGCCGGGGGCGCTCC 360

Qy    2001 GGACTCTTTCCCGATGACGTGGTGCAGCCAGCTGCTGCCCCGACCTCTCCTTTCCCTG 2060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGACTCTTTCCCGATGACGTGGTGCAGCCAGCTGCTGCCCCCGACCTCTCCTTTCCCTG 419

Qy    2061 GAAAAAGAAAACAGCTGGCAACGCAAGAGTAAGCTGGGGCCAGCTCAGGAGGTGAGGAAG 2120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GAAAAAGAAAACAGCTGGCAACGCAAGAGTAAGCTGGGGCCAGCTCAGGAGGTGAGGAAG 479

Qy    4121 ACAGAAGAGGGTGAAGTGATACAGGCTAACTTGGAGACTGAGAAGAAAGAGCAGGTT 4180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACAGAAGAGGGTGAAGTGATACAGGCTAACTTGGAGACTGAGAAGAAAGAGCAGGTT 539

Qy    4181 CTTCGGCTGTTGTCCACTTCCTGTCTGGTGCCAGGCTAACTCTTTCTTTTA 4240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 CTTCGGCTGTTGTCCACTTCCTGTCTGGTGCCAGGCTAACTCTTTCTTTTA 598

Qy    4241 GTATCTCCTGACTTTTTGCCATTTGTGAGACTGTAATGTTACACTTCTTAACTTTTTATCT 4300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     599 GTATCTCCTGACTTTTTGCCATTGTGAGACTGTAATGTTACACTTCTTAACTTTTTATCT 658

Qy    4301 TAGTTCAGTGTCTCCATAGAGGATGTTAATAAATAAAGTTTTTTTTTGAAAAAAAAA 4360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     659 TAGTTCAGTGTCTCCATAGAGGATGTTAATAAATAAAGTTTTTTTTTGAAAAAAAAA 717

Qy    4361 AAAAAAAAAAAAAA 4374
           ||||||||||||||
Db     717 AAGAGAAACAAACAA 730


RESULT 5
BB072116
LOCUS       BB072116                677 bp    mRNA    linear   EST 18-OCT-2002
DEFINITION  BB072116 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030008I18 3', mRNA sequence.
ACCESSION   BB072116
VERSION     BB072116.2  GI:16260140
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hamaguchi,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Itoh,M., Kawai,J., Kondo,S., Kono
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohara,O.,
            Okazaki,Y., Okido,T., Saito,R., Sasaki,D., Sawaki,K., Sato,H., Shiraki
            ,T., Shibata,K., Shinagawa,A., Shiroki,T., Tagami,M., Yoshiki,A.,
            Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyo,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 19, 2002 this sequence version replaced gi:18870114.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research, (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of Cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            Iwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Ohsumi
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
```

Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030008I18"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      148 a    176 c    186 g    165 t      2 others
ORIGIN
```

```
  Query Match             15.0%;  Score 657.6;  DB 10;  Length 677;
  Best Local Similarity   99.0%;  Pred. No. 1e-104;
  Matches  671;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy    3674 AGTTTCCACTTGGACCAGATGAAGACACTCTGTTCCTACAGTTATG-TGAAGTCCTGACC 3733
           ||||||||| |||||||| |||||||||||||||||||||||||| ||| |||||||||||
Db       1 AGTTTCCACTTGGACCAGATGAAGACACTCTGTTCCTACAGTTATG-TGAAGTCCTGACC 59

Qy    3734 GTGCAGTGCAAGGGCAGATCCACCCTGGAGCTGTCCTTGAAGAATGAGCAGCTGATACTG 3793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 GTGCAGTGCAAGGGCAGATCCACCCTGGAGCTGTCCTTGAAGAATGAGCAGCTGATACTG 119

Qy    3794 CACACAGCCTGGGCGAGGGCCATCAAGGCCATGGTGGATCTATTTCTGAGTGAACTCAGG 3853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 CACACAGCCTGGGCGAGGGCCATCAAGGCCATGGTGGATCTATTTCTGAGTGAACTCAGG 179

Qy    3854 AAGGACTCCGGCTATGTCATCGGCCTGCGCAGCTACATCACCGATGACAATAGCCTCCTC 3913
           |||||||||| |||||||||||||||||||| ||||||||||||||||||||||||||||
Db     180 AAGGACTCCGGTTATGTCATCGGCCTGCGCAGNTACATCACCGATGACAATAGCCTCCTC 239

Qy    3914 AGTTTCCA[illegible]CTTACTGGCAGTGACCGCTCTGGAACCAGGCTGG 3973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 AGTTTCCA[illegible]CTTACTGGCAGTGACCGCTCTGGAACCAGGCTGG 299

Qy    3974 CAGTTCGGTTCGTGCCCCCCCCCCCTCCGGACTCTTTCCCGATGACGTGGTGCAGCCAGCT 4033
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 CAGTTCGGTTCGTGCCCCCCCCCCCTCCGGACTCTTTCCCGATGACGTGGTGCAGCCAGNT 359

Qy    4034 GCTGCCCCCGACCTCTCCTTTTCCCTGGGAAAGAGAAACAGCTGGCAACGCAAGAGTAAG 4093
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 GCTGCCCCCGACCTCTCCTTTTCCCTGGGAAAGAGAAACAGCTGGCAACGCAAGAGTAAG 419

Qy    4094 CTGGGGCCAGCTCAGGAGGTGAGGAAGACAGAAGAGGTGAAGTGATACAGGCCTAACTTG 4153
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 CTGGGGCCAGCTCAGGAGGTGAGGAAGACAGAAGAGGTGAAGTGATACAGGCCTAACTTG 479

Qy    4154 GAGACTGAGAAGGAAAGAGCAGGGTTGCTTCGGGTGTTGTCCACTTCCTGTCCTGGTGGC 4213
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 GAGACTGAGAAGGAAAGAGCAGGGTTGCTTCGGGTGTTGTCCACTTCCTGTCCTGGTGGC 539

Qy    4214 CAGGGCTCAATGTGTTCCTGTCCTTTACCATCTCCTGACTTTTTGCCATTTGTGAGACTG 4273
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 CAGGGCTCAATGTGTTCCTGTCCTTTACCATCTCCTGACTTTTTGCCATTTGTGAGACTG 599

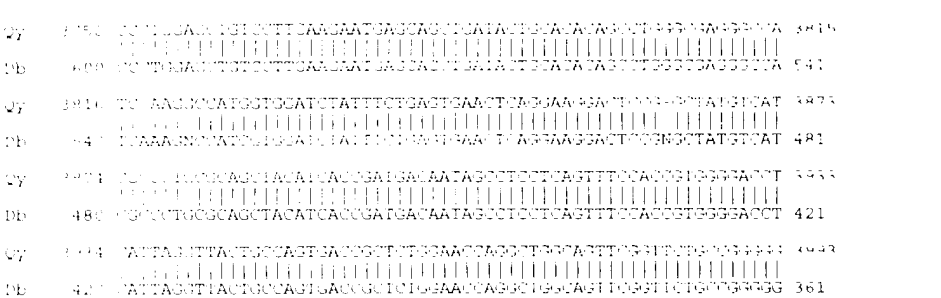
Qy    4274 TAAGTCACACCCTCTAACTCTGGTACTTAGTTCAGTGTCTCCATAGAGGATGCTTAATAA 4333
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 TAAGTCACACCCTCTAACTCTGGTACTTAGTTCAGTGTCTCCATAGAGGATGCTTAATAA 659

Qy    4334 ATAACCTTGGTTTTCCTG 4351
           ||||||||||||||||||
Db     660 ATAACCTTGGTTTTCCTG 677
```

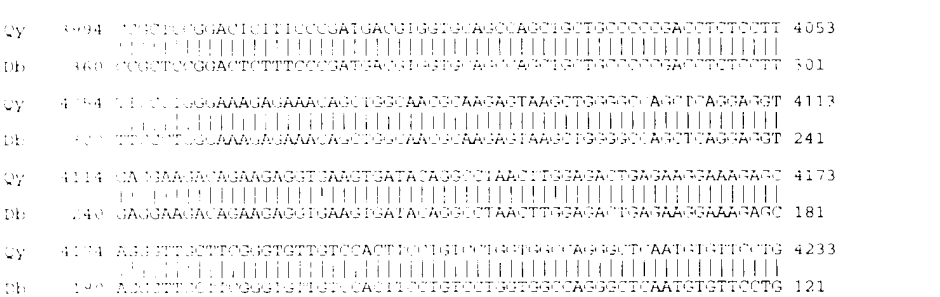
RESULT 6
BI851062/c
LOCUS BI851062 610 bp mRNA linear EST 10-OCT-2001
DEFINITION image:qc_12_2000/smd407bdff41.x1 NCI_CGAP_SI1 Mus musculus cDNA clone IMAGE:4353317 3', mRNA sequence.
ACCESSION BI851062
VERSION BI851062.1 GI:16004549
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 610)
AUTHORS Kale,P.I., Harsch,T.J., Forta,P.A., Nelson,D.O., Sanders,C.G. and Prange,C.K.
TITLE The I.M.A.G.E. Consortium quality control effort: clone resequencing for verification
JOURNAL Unpublished (2001)
COMMENT Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been produced as part of the I.M.A.G.E. Consortium quality control effort. High quality sequence is defined as having 100 or more base pairs with a phred quality value of 20 or greater, where a sliding window of 4 base pairs marks the beginning and end of the sequence. For information on obtaining this clone, please contact info@image.llnl.gov
Plate: LIAM0986 row: b column: 6
Seq primer: -21m13
High quality sequence stop: 610.

```
FEATURES             Location/Qualifiers
     source          1..610
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4353317"
                     /clone_lib="NCI_CGAP_SI1"
                     /sex="male"
                     /dev_stage="5 months"
                     /lab_host="DH10B (T1-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6 (Life
                     Technologies); Site_1: Not I; Site_2: Sal I; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     1.63 Kb. Library constructed by Life Technologies."
BASE COUNT      147 a    166 c    163 g    132 t      2 others
ORIGIN
```

```
  Query Match             13.1%;  Score 574;  DB 12;  Length 610;
  Best Local Similarity   99.5%;  Pred. No. 3.8e-90;
  Matches  596;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;
```



RESULT 7
BF582477
LOCUS BF582477 958 bp mRNA linear EST 12 DEC 2000
DEFINITION [illegible] NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4208733 5', mRNA sequence.
ACCESSION BF582477
VERSION BF582477.1 GI:11556199
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 958)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9773 row: l column: 10
High quality sequence start: 5
High quality sequence stop: 614.
FEATURES Location/Qualifiers
source 1..958
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4208794"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: colon. Vector: pCMV SPORT6; Site_1: Not I; Site_2: Sal I; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.6 Kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
BASE COUNT 248 a 297 c 266 g 157 t
ORIGIN



RESULT 8
BB624737
LOCUS BB624737 658 bp mRNA linear EST 20-[illegible]-2002
DEFINITION BB624737 RIKEN full-length enriched [illegible] Mus musculus cDNA clone 9030009I18 5', mRNA sequence.
ACCESSION BB624737
VERSION BB624737.1 GI:16463169
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

EFERENCE 1 (bases 1 to 653)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Ohzaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sugabo,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse FSTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
OMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku,,Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome res@gsc.riken.go.jp
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,F., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384 format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,Y., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1..653
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9030008I18"
/clone_lib="RIKEN full-length enriched, adult male colon"
/sex="male"
/tissue_type="colon"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI, Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"

BASE COUNT 127 a 215 c 187 g 123 t 1 others
ORIGIN

Query Match 12.6%; Score 582.2; DB 10; Length 653;
Best Local Similarity 97.1%; Pred. No. 2.3e-86;
Matches 573; Conservative 0; Mismatches 14; Indels 3; Gaps 1;

Qy 860 CCCGCCCAGCTGGCTACTCTGCTGGAGAGGGCGGAA-GGCCAGCAGTTTTTGACG 916
Db 62 CCCGCACAGCGGGGAGCCCTGGCGGAAAGTGTCAAATATGGCCAGCAGGCTTGCTGACG 121

Qy 917 GGGAGCATCACAGAGTCCCTGCCAGCTGAGGTCCCCGCCCGGCCCAGCCTGACTCTCCCT 976
Db 122 GGGAGCATCACAGAGTCCCTGCCAGCTGAGGTCCCCGCCCGGCCCAGCCTGACTCTCCCT 181

Qy 977 CCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCAGAAGCCATTT 1036
Db 182 CCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCAGAAGCCATTT 241

Qy 1037 CTGCCTCGACCAGGGCAGCCACTGGACGAGCCCCTGACGCGGTTAGATGGCGAGAACCCT 1096
Db 242 CTGCCTCGACCAGGGCAGCCACTGGACGAGCCCCTGACGCGGTTAGATGGCGAGAACCCT 301

Qy 1097 CAGCAGGCTCTGGAGATCAACAGGGTGATGCTGCGGCTCCTCTGGGGAAGGATCTTCGCAG 1156
Db 302 CAGCAGGCTCTGGAGATCAACAGGGTGATGCTGCGGCTCCTCTGGGGAAGGATCTTCGCAG 361

Qy 1157 TCCTGGCAAGAGCAGACCATGGGCACGTTCCTCGTGCAGCAGGCCAGCCACCGCCGGGA 1216
Db 362 TCCTGGCAAGAGCAGACCATGGGCACGTTCCTCGTGCAGCAGGCCAGCCACCGCCGGGA 421

Qy 1217 CTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAGCTGTGGGGCAACCCAGATGAGCAA 1276
Db 422 CTCCGAGATGAGCTCTTCAGCCAGCTGGTGGCCCAGCTGTGGGGCAACCCAGATGAGCAA 481

Qy 1277 CAGAATCAGCGTGGCTGGGCCCTAATGGTGATCCTGCTCAGCTCCTTTGCTCCCACACCT 1336
Db 482 CAGAATCAGCGTGGCTGGGCCCTAATGGTGATCCTGCTCAGCTCCTTTGCTCCCACACCT 541

Qy 1337 GCCCTGGAGAAGCCACTGCTCAAATTTGTATCTGACCAGGCTCCCAGTGGCATGGCAGCC 1396
Db 542 GCCCTGGAGAAGCCACTGCTCAAATTTGTATCTGACCAGGCTCCCAGTGGCATGGCAGCC 601

Qy 1397 CTGTGCCAGCACAAGCTGTTAGGTGCCCTGGAGCANACACCGCTGGCTCC 1446
Db 602 CTGTGCCAGCACAAGCTGTTAGGTGCCCTGGAGCANACACCGCTGGCTCC 651

RESULT 9
BF585148
LOCUS BF585148 929 bp mRNA linear EST 12-DEC-2000
DEFINITION 602101892F2 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4225198 5',
mRNA sequence.
ACCESSION BF585148
VERSION BF585148.1 GI:11658866
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```
            Plate: LLAM9816  row: g  column: 23
            High quality sequence stop: 715.
FEATURES             Location/Qualifiers
     source          1..929
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4225198"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally  Primer: Oligo dT
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library "
BASE COUNT      296 a    285 c    278 g    160 t
ORIGIN

  Query Match             12.3%;  Score 536;  DB 12;  Length 929;
  Best Local Similarity   99.8%;  Pred. No. 1.4e-83;
  Matches  547;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy     [illegible]
Db      10 [illegible]

Qy     [illegible]
Db      70 [illegible]

Qy     [illegible]
Db     130 [illegible]

Qy     [illegible]
Db     190 [illegible]

Qy     [illegible]
Db     250 [illegible]

Qy     [illegible]
Db     310 [illegible]

Qy     [illegible]
Db     370 [illegible]

Qy     [illegible]
Db     430 [illegible]

Qy     [illegible]
Db     490 [illegible]

Qy     4194 CTCAGGAG 4201
            ||||||||
Db      560 CTCAGGAG 567


RESULT 13
BM926425
LOCUS       BM926425                1088 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6100221 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5590384
            5', mRNA sequence.
ACCESSION   BM926425
VERSION     BM926425.1  GI:19356804
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1088)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection  MGC
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I M A G E  Consortium (LLNL
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I M A G E  Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12364  row: j  column: 17
            High quality sequence stop: 567.
FEATURES             Location/Qualifiers
     source          1..1088
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5590384"
                     /clone_lib="NIH_MGC_125"
                     /lab_host="DH10B"
                     /note="Tissue: ovary; Vector: pCMV SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 20 to
                     49 yo. Library is oligo dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1.5 kb. Library is
                     normalized and enriched for full length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
BASE COUNT      182 a    374 c    345 g    187 t
ORIGIN

  Query Match             11.5%;  Score 504.4;  DB 14;  Length 1088;
  Best Local Similarity   73.5%;  Pred. No. 4.3e-78;
  Matches  683;  Conservative    0;  Mismatches  246;  Indels   13;  Gaps    4;

Qy     [illegible]
Db      21 [illegible]

Qy     [illegible]
Db      81 [illegible]

Qy     [illegible]
Db     141 [illegible]

Qy     [illegible]
Db     201 [illegible]

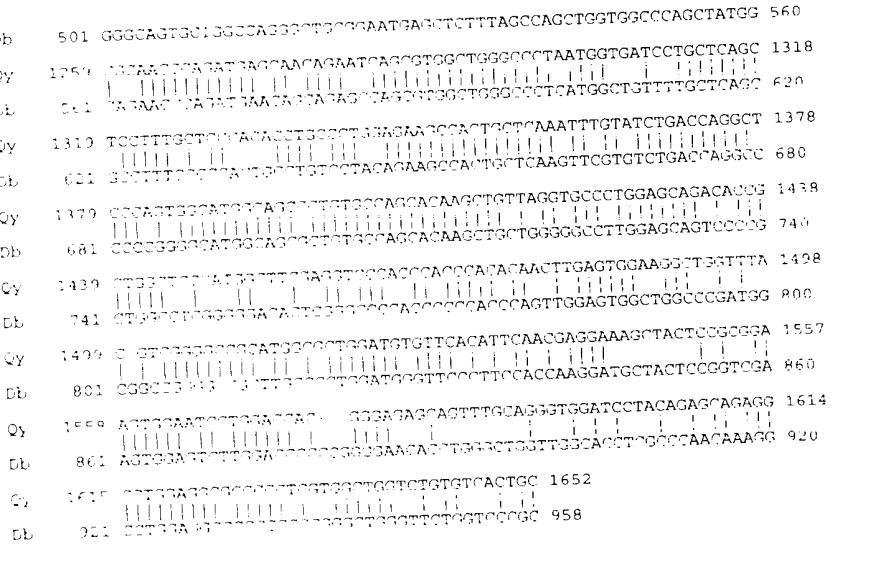
Qy     [illegible]
Db     261 [illegible]

Qy     [illegible]
Db     321 [illegible]

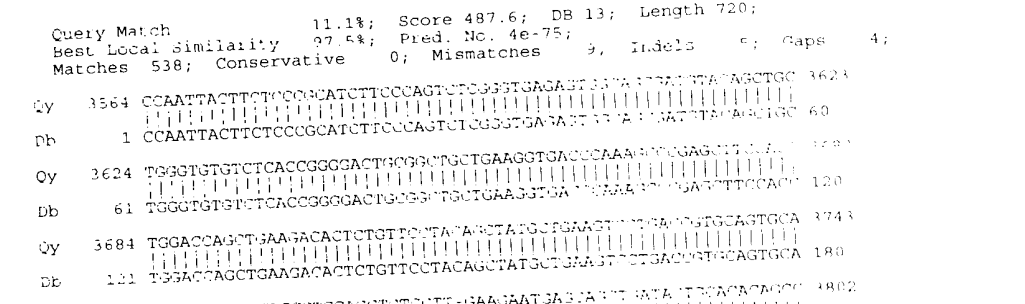
Qy     [illegible]
Db     381 [illegible]

Qy     [illegible]
Db     441 [illegible]

Qy     [illegible]
```



RESULT 11
BG964584
LOCUS BG964584 710 bp mRNA linear EST 12-JUN-2001
DEFINITION 6029+1563F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986830 5',
mRNA sequence.
ACCESSION BG964584
VERSION BG964584.1 GI:14352001
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 720)
AUTHORS NIH MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10996 row: n column: 15
High quality sequence start: 4
High quality sequence stop: 710.
FEATURES Location/Qualifiers
source 1..720
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4986830"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 162 a 219 c 195 g 144 t
ORIGIN



Query Match 11.1%; Score 487.6; DB 13; Length 720;
Best Local Similarity 97.6%; Pred. No. 4e-75;
Matches 538; Conservative 0; Mismatches 9; Indels 5; Gaps 4;

RESULT 12
AI891960
LOCUS AI891960 505 bp mRNA linear EST 27-JUN-1999
DEFINITION ul57a11.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
IMAGE:2123132 5' similar to SP:CO2A6_CHICK MYOSIN XV ;, mRNA
sequence.
ACCESSION AI891960
VERSION AI891960.1 GI:5597862
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 505)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,M., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: ul57a11.x1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

```
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information
            MGI:999328
            Possible reversed clone; similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1..505
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2123132"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG), Site_2: DraIII (CACGATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCGTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site (CACTGTGTG, 3' site CACCATGTG).  XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers for
                     sequencing: 5' end primer GTTCTGGTCTAAAACTGCG and 3' end
                     primer CGACGTGCAGCTCGAGCACA."
BASE COUNT       96 a    165 c    154 g     90 t
ORIGIN

  Query Match             11.1%;  Score 484;  DB 9;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-74;
  Matches  484;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      787 GTGTGAGCCAGGGGGGAGGCCTGGGGGAAAGTGTCAAATATGGACCTGGG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       41 GTGTGAGCCAGGGGGGAGGCCTGGGGGAAAGTGTCAAATATGGACCTGGG 90

Qy      841 TGCTTAGAGATCCCCGGCCAGTGGCTACTCTGCTGGAGAGGGGGGAAGGCCACCAGGC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       91 TGCTTAGAGATCCCCGGCCAGTGGCTACTCTGCTGGAGAGGGGGGAAGGCCACCAGGC 150

Qy      901 CTTGTGGAGGGGAGCATCACAGAGTCCCTGCCAGCTGAGGTCCCCGCCGGGCCAGCCT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      151 CTTGTGGAGGGGAGCATCACAGAGTCCCTGCCAGCTGAGGTCCCCGCCGGGCCAGCCT 210

Qy      961 GACTCTCCTCCAGACATTGACCAGTTTCCCTTCTCCATTTTTGTATCCACAGCTTTCA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      211 GACTCTCCTCCAGACATTGACCAGTTTCCCTTCTCCATTTTTGTATCCACAGCTTTCA 270

Qy     1021 GAACTCATTTCTGCCTCGACCAGGGCAGCCACTGGACCAAGCCCTGAAGAGGGTTACAAGG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      271 GAACTCATTTCTGCCTCGACCAGGGCAGCCACTGGACCAAGCCCTGAAGAGGGTTACAAGG 330

Qy     1081 TGAAAACCTCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCTGGGGCTCCTGGGGGAAGG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      331 TGAAAACCTCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCTGGGGCTCCTGGGGGAAGG 390

Qy     1141 ATCAGCCGGGGACTCCGAGATGAGCTCTTTAGCTACGTGTCGACCAGCTGTGGGCGCAACCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      391 ATCAGCCGGGGACTCCGAGATGAGCTCTTTAGCTACGTGTCGACCAGCTGTGGGCGCAACCC 450

Qy     1201 AGATCTGT 1210
            ||||||||
Db      451 AGATCTGT 458
```

```
RESULT 13
BI294877/c
LOCUS       BI294877                 698 bp    mRNA    linear   EST 10-JUL-2002
DEFINITION  UI-R-DK0-cod-b-10-0-UI.s1 UI-R-DK0-Rattus norvegicus cDNA clone
            UI-R-DK0-cod-b-10-0-UI 3', mRNA sequence.
ACCESSION   BI294877
VERSION     BI294877.1  GI:14957760
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: mbsoares-blue@uiowa.edu
            The sequence contained in oligo dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a genuine poly A
            tail.  The sequence tag present in the cDNA between the Not I site
            and the oligo-dT track served to identify it as a clone from the
            normalized rat heart pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..698
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-DK0-cod-b-10-0-UI"
                     /clone_lib="UI-R-DK0"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; The UI_R_DK0
                     library is a subtracted library derived from a mixture of
                     five individually tagged normalized rat libraries:
                     brain nRBP (20%), heart nRBP (20%), kidney nRBP (20%),
                     aorta-nPAP (20%), and placenta nPPP (20%). Each original
                     library was constructed from a mixture of equal amounts of
                     RNA from seven different developmental time-points:
                     embryonic day 12, embryonic day 17, embryonic day 21,
                     adult day 1, adult day 12, adult day 75, and adult day
                     200. (Exception: the aorta pool does not contain embryo day
                     17 RNA and the placenta pool contains only the three
                     embryonic stages). Each library was normalized
                     individually according to the procedure described by
                     Bonaldo, Lennon & Soares (Genome Research Genome 6:
                     791-806, 1996). For construction of the DK0 subtracted
                     library, plasmid DNA from each of the five individually
                     tagged normalized libraries was mixed in the proportions
                     specified above and electroporated into competent bacteria
                     for production of single stranded circular DNA
                     representing the pool of libraries. Single stranded
                     circular DNA representing these five normalized libraries
                     was then used as a tracer in a subtractive hybridization
                     with a driver (PCR amplified inserts from a plasmid DNA
                     template preparation) comprising: a) a set of about 1,000
                     arrayed clones from each of the five non-normalized
                     libraries of brain (CT0s), heart (CT3s), kidney (CT5s),
```

aorta (CW0s), and placenta (CX0s). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CT0), heart (CS0), kidney (CU0), aorta (CW0), and placenta (CX0). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population."
TAG_LIB=UI-R-DK0
TAG_TISSUE=rat heart pool
TAG_SEQ=ATAAGATAAC"

BASE COUNT 150 a 203 c 187 g 158 t
ORIGIN

Query Match 10.9%; Score 477; DB 13; Length 698;
Best Local Similarity 86.7%; Pred. No. 2.8e-73;
Matches 599; Conservative 0; Mismatches 75; Indels 17; Gaps 6;

```
Qy    3696 AGACACTCTGTTCCTACAGCTATGCTGAAGTCCTGACCGTGCAGTGCAGGGGCAGATCCA 3755
           ||  |  || |  | |||||||||| || ||  || | |||||||| || |||  |||
Db     689 AGCCTGGCTCTGTCCACAGCTATGCCGAGGTGTTGGCTGTGCAGTGTAGGAGCACCTCCA 630

Qy    3756 CCCTGGAGCTGTCCTTGAAGAATGAGCAGCTGATACTGCACACAGCCTGGGCGAGGGCCA 3815
           ||||||||||||  |||||| |||||||||||| |||||||  |||||||||| || ||
Db     629 CCCTGGAGCTGTCACTGAAGAGTGAGCAGCTGATCCTGCACACGACCTGGGCGAGGACCA 570

Qy    3816 TCAAGGCCATGGTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTCATCG 3875
           | |||||||||||||| |  || |||||||| ||| ||||||||||||| || |||||||
Db     569 TGAAGGCCATGGTGGAACAGTTCCTGAGTGAGCTCAAGAAGGACTCCGGTTACGTCATCG 510

Qy    3876 CCCTGTGCAGCTACATCACCGATGACAACAGTCTCCTCAGTTTCCACCGTGGGGACCTCA 3935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 CCCTGCGCAGCTACATCACCGATGACAACAGTCTCCTCAGTTTCCACCGTGGGGACCTCA 450

Qy    3936 TTAGGTTACTCTCAATGGACTTTCTGCAACCAGGCTGGCAGTTCGGTTCTGCCGGGGGCC 3995
           |||||  |||||| ||| |||||||||||||||||||||||||||||| ||||||||||
Db     449 TTAGGCTGCTGCAGTGGTTTCTGCAACCAGGCTGGCAGTTCGGCTCTGCCGGGGGCC 390

Qy    3996 GATCAGGACTCTTTTCCGATGACGTTGTGCAGCCAGCTGCTGCCCCGGACCTCTCCTTTT 4055
           |||||||||||||||| |||| ||||||||||||||||||||||||||| |||||||| |
Db     389 GCTCCGGACTCTTTCCGTCTGACATGGTGCAGCCAGCTGCTGCCCCGGACCTCTCCTTCT 330

Qy    4056 CCCTGGGAAAGAGAAACAGCTGGCAACGCAAGAGTAAGCTG------GGGCCAGCTCAGG 4109
           | |||||||||||||||||||||||| ||||||||||||||      ||||||||||||
Db     329 CGCTGGGAAAGAGAAACAGCTGGCAGCGCAAGAGTAACAGAGTAACAGGGCCAGCTCAGG 270

Qy    4110 AGGTGAGGAAGACAGA--AGAGGTGAAGTGAT-ACAGGCCTAACTTGGAGACTGAGAAGG 4166
           |||||||||||||||||  |||||||||||||| | |||||||||| |||| || ||||||
Db     269 AGGTGAGGAAGACAGAAGAGAGGTGAAGTGATAAAAGGCCTAACTCGGAGTCTCAGAAGG 210

Qy    4167 AAAGAGCAGGGTTGCTTGGGGTGTTGTCCACTTCC--TGTCCTGGTGGCCAGGGCTCAAT 4224
           |||||||  |||||||||||||| ||  |||||||  || | |||||||| ||||||||||
Db     209 AAAGAGCGGGGTTGCTTGGGGTGCTGCTCACTTCCGGTGACGTGGTGGCCAGTGCTCAAT 150

Qy    4225 GTGTTCCTGTCCTTTACCAT ---CTCCTGACTTTTTGCCATTTGTGAGACTGTAAGTCA 4280
           ||||||||  |||||| || |  ||||||||||||||||||||||||||||||||||||
Db     149 GTGTTCCTATCCTTTGCCGTCTGCCTCCTGACTTTTTGCCA--TGTGAGACTGTAAGCCA 92

Qy    4281 CAGCCTCTAACTCTGGTACTTAGTTCAGTGTCTCCATAGAGGATGCTTAATAAATAACCT 4340
           | |||||| |||||||||||| || |||| ||| ||||||||||||||||||||||||||
Db     91 CGGCTTCTGACTCTGGTACTTAGGTCGGTGTCTTCATGGAGGATGCTTAATAAATAACCT 32

Qy    4341 TGGTTTTCCTGGAAAAAAAAAAAAAAAAAA 4371
           ||| |||||||| |||||||||||||||||
Db     31 TGGCTTTCCTGGCAAAAAAAAAAAAAAAAA 1
```

RESULT 14
BM791149
LOCUS BM791149 603 bp mRNA linear EST 05-MAR-2002
DEFINITION K-EST0071111 S21SNU520 Homo sapiens cDNA clone S21SNU520-19-F03 5', mRNA sequence.
ACCESSION BM791149
VERSION BM791149.1 GI:19139381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 603)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 10 row: F column: 03
High quality sequence stop: 603.

FEATURES Location/Qualifiers
source 1..603
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNU520-19-F03"
/clone_lib="S21SNU520"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F'"
/note="Organ: Stomach, Vector: pYT19RP1; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoPI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 109 a 186 c 187 g 121 t
ORIGIN

Query Match 9.8%; Score 430.2; DB 14; Length 603;
Best Local Similarity 82.1%; Pred. No. 4.1e-65;
Matches 495; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

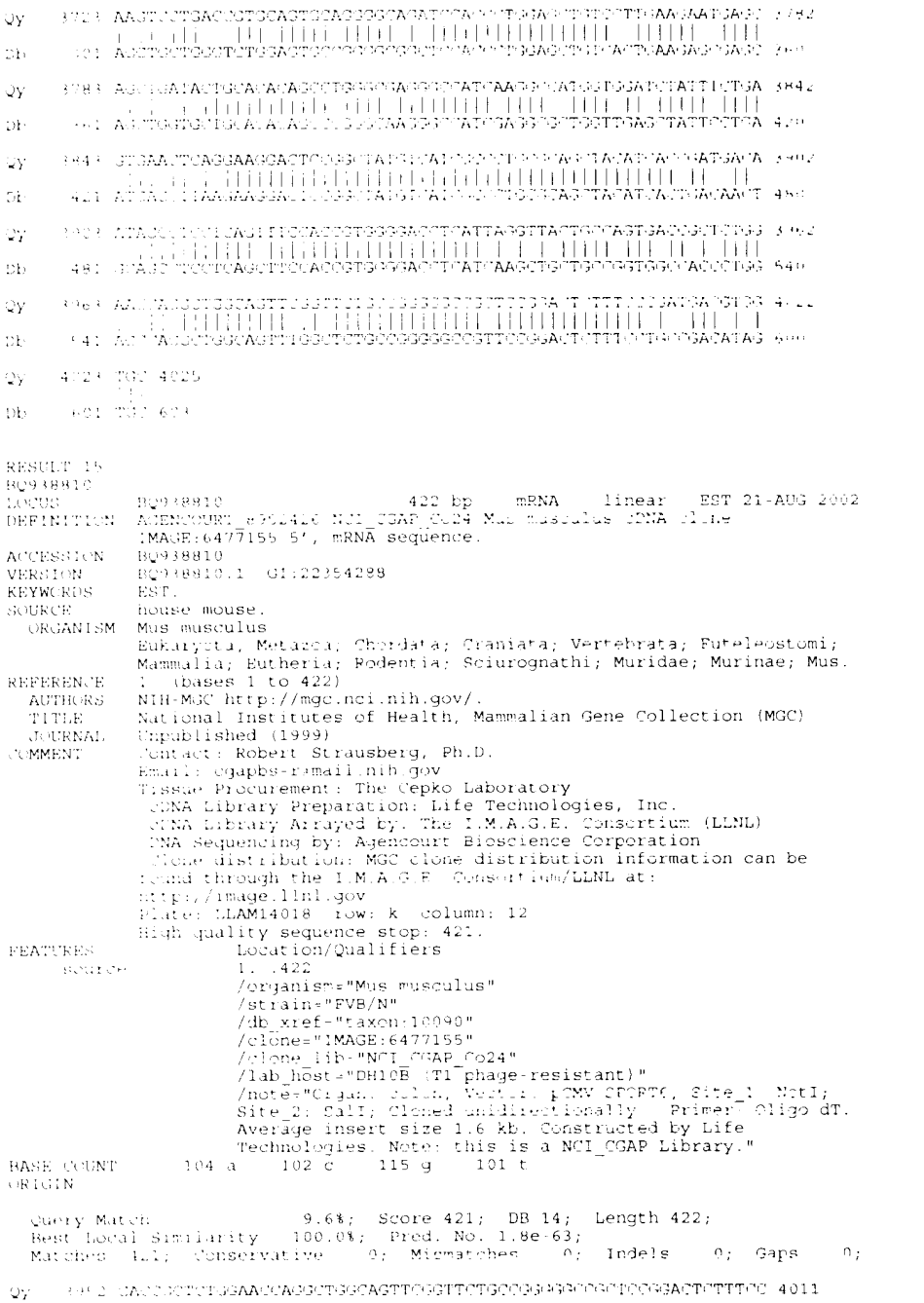
```
Qy    3423 AGTCCTGCACCCGGATCTCACAGGATGAGCGGCACAAAATGAAAGGCCTTCTGGGAGACT 3482
           |||||||  |||||| || ||| |||||||||   |||||||||| || |||||||  ||
Db       1 AGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGGCT 60

Qy    3483 TGGAGGTGAGTCTGGAGACCCTTGACATTGTTGAAGACAGCATCAAAAAACGCATCGTGG 3542
           ||||||||   |||||  | ||   ||  |||||||| |||| || ||||||||||||||
Db      61 TGGAGGTGGACCTGGATTCTCTCACCACCACGAAGACAGGGTCAAGAAGCGCATTGTGG 120

Qy    3543 TCGCTGCTCGGGACAACTGGGCCAATTACTTCTCCCGCTTCTTCCTGTCTGTCGGGAGA 3602
           | ||||||||||||||||||||||||||||||||||||||||| |||| || |||||||| ||||
Db     121 TGGCCGCTCGGGACAACTGGGCCAATTACTTCTCCCGCTTCTTCCTGTCTGTCGGGAGA 180

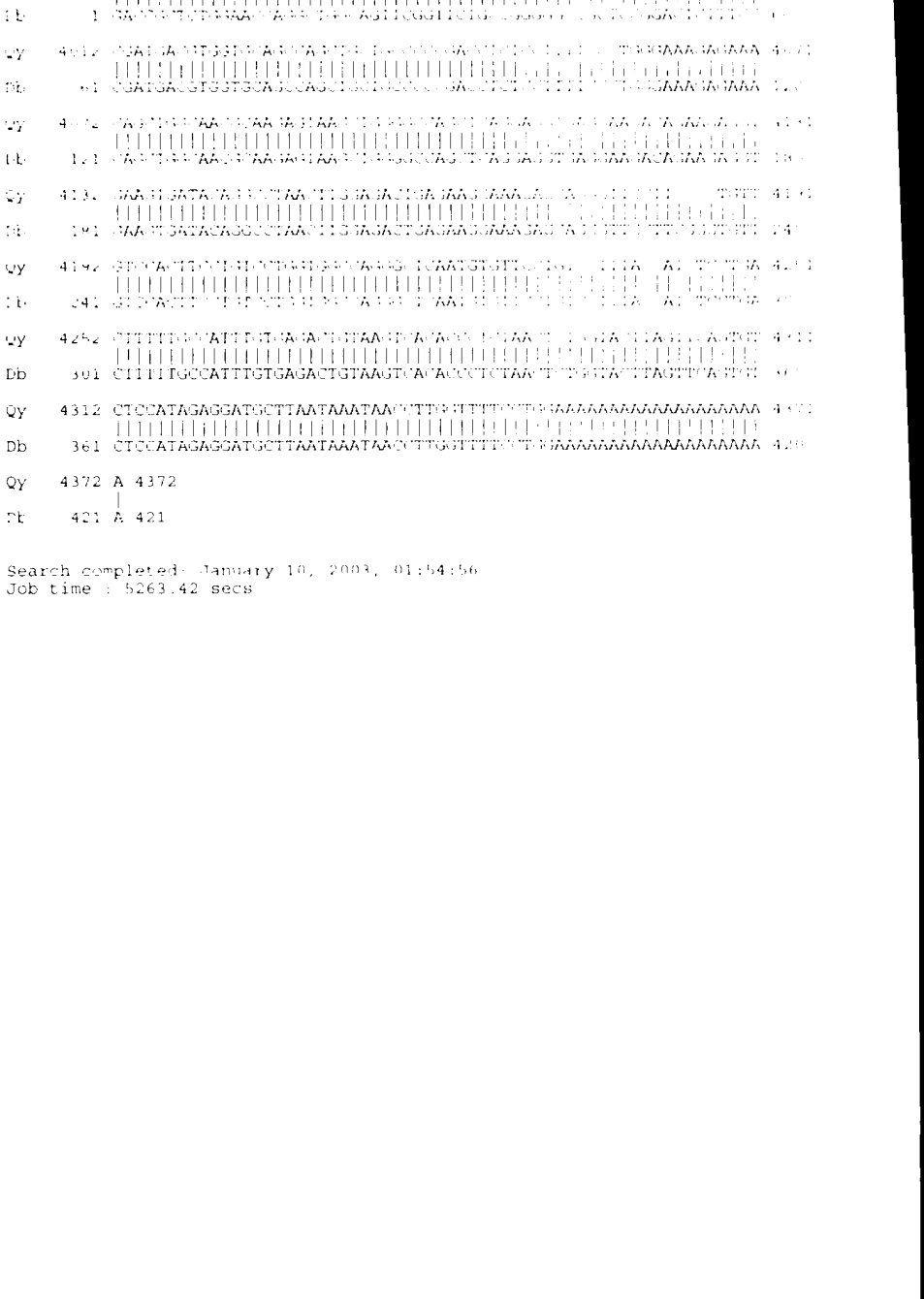
Qy    3603 GTGGCAGCGATGTACAGCTGCTGGGTGTGTCTCACCGGGGACTGCAGGCTGCTGAAGGTGA 3662
           |||||||||| || |||||| | |  ||||| |||||  |||||  |||||||||||||||
Db     181 GTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTGAAGGTGA 240

Qy    3663 CCCAAAGCCCGAGCTTCCAGCTGGACCAGCTGAAGACACTCTGTTCCTACAGCTATGCTG 3722
           ||||| ||||  || ||| ||  |||||||||||||  |||||  || |||||||  ||| |
Db     241 CCCAAGGCCCCGGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGCGG 300
```



RESULT 15
BQ938810
LOCUS BQ938810 422 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8952410 NCI_CGAP_Co24 Mus musculus cDNA clone
IMAGE:6477155 5', mRNA sequence.
ACCESSION BQ938810
VERSION BQ938810.1 GI:22954298
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 422)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14018 row: k column: 12
High quality sequence stop: 421.
FEATURES Location/Qualifiers
source 1..422
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6477155"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 104 a 102 c 115 g 101 t
ORIGIN

Query Match 9.6%; Score 421; DB 14; Length 422;
Best Local Similarity 100.0%; Pred. No. 1.8e-63;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Search completed: January 10, 2003, 01:54:56
Job time : 5263.42 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd

SM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:10:54 ; Search time 124.282 Seconds
(without alignments)
10795.718 Million cell updates/sec

Title: US-09-803-126-3
Perfect score: 4375
Sequence: 1 [illegible] aaaaaaaaaaaaaaaaaaaa 4375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters. 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMP.seq:*
2: /cgn2_6/ptodata/1/ina/6A_COMP.seq:*
3: /cgn2_6/ptodata/1/ina/6B_COMP.seq:*
4: /cgn2_6/ptodata/1/ina/6P_COMP.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMP.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
C	1	55	1.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	46.8	1.1	364	4	US-09-905-223-162	Sequence 162, App
	3	44.8	1.0	776	4	US-09-535-008-36	Sequence 36, Appl
C	4	43	1.0	1357	4	US-09-668-680-3	Sequence 3, Appli
C	5	42.4	1.0	71988	4	US-09-443-501A-2	Sequence 2, Appli
	6	41.6	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	7	41	0.9	3489	2	US-08-728-323A-1	Sequence 1, Appli
	8	41	0.9	3489	4	US-09-298-568-1	Sequence 1, Appli
C	9	41	0.9	32207	2	US-08-770-379-20	Sequence 20, Appl
C	10	41	0.9	32207	4	US-08-757-669A-20	Sequence 20, Appl
C	11	41	0.9	32207	4	US-09-230-371A-20	Sequence 20, Appl
C	12	40.8	0.9	68750	2	US-08-335-409-1	Sequence 1, Appli
C	13	40.8	0.9	68750	4	US-09-568-102-1	Sequence 1, Appli
C	14	40.8	0.9	68750	4	US-09-567-969-1	Sequence 1, Appli
C	15	40.8	0.9	68750	4	US-09-568-480-1	Sequence 1, Appli
C	16	40.8	0.9	68750	4	US-09-568-496-1	Sequence 1, Appli
C	17	40.8	0.9	68750	4	US-09-568-472-1	Sequence 1, Appli
C	18	40.8	0.9	68750	4	US-09-567-899-1	Sequence 1, Appli
	19	40	0.9	2074	4	US-08-630-915A-19	Sequence 19, Appl
C	20	40	0.9	3251	4	US-09-080-199B-6	Sequence 6, Appli
	21	39.8	0.9	1160	1	US-08-161-406-1	Sequence 1, Appli
	22	39.6	0.9	2984	5	PCT-US93-00893-1	Sequence 1, Appli
C	23	39.6	0.9	2984	5	PCT-US93-00893-2	Sequence 2, Appli
C	24	39.4	0.9	51259	3	US-08-781-891-209	Sequence 209, App
	25	39	0.9	3059	4	US-09-276-531-107	Sequence 107, App
	26	38.8	0.9	477	4	US-09-135-994-1	Sequence 1, Appli
C	27	38.6	0.9	16442	3	US-08-781-891-208	Sequence 208, App
C	28	38.2	0.9	2793	1	US-08-209-747-1	Sequence 1, Appli
C	29	38.2	0.9	2793	1	US-08-458-298-1	Sequence 1, Appli
C	30	38	0.9	1926	4	US-09-249-585A-2	Sequence 2, Appli
C	31	38	0.9	2580	3	US-08-359-863-2	Sequence 2, Appli
C	32	38	0.9	2580	4	US-09-359-081-2	Sequence 1, Appli
	33	38	0.9	5452	2	US-09-130-114-1	Sequence 1, Appli
C	34	38	0.9	9600	4	US-08-910-647-1	Sequence 1, Appli
C	35	38	0.9	9600	4	US-09-620-925-1	Sequence 1, Appli
C	36	38	0.9	10596	1	US-07-884-811-15	Sequence 15, Appl
C	37	38	0.9	10596	1	US-07-885-971-15	Sequence 15, Appl
C	38	38	0.9	10596	1	US-08-087-783A-15	Sequence 15, Appl
C	39	38	0.9	10596	1	US-08-194-088B-15	Sequence 15, Appl
C	40	38	0.9	10596	2	US-08-194-087-15	Sequence 15, Appl
C	41	38	0.9	10596	5	PCT-US93-04648-15	Sequence 15, Appl
	42	37.8	0.9	4502	4	US-09-404,650-12	Sequence 12, Appl
	43	37.6	0.9	1962	2	US-08-919-421A-1	Sequence 1, Appli
	44	37.6	0.9	1962	4	US-09-342-819-1	Sequence 1, Appli
	45	37.6	0.9	4217	4	US-09-232-200-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1900 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTOgpt-F1s
US-08-232-463-14

Query Match 1.3%; Score 55; DB 1; Length 7218;

```
  Best Local Similarity   4.0%;  Pred. No  0.0015;
  Matches   17;  Conservative  223;  Mismatches  158;  Indels    0;  Gaps    0;

Qy     2381 CAAGAGAAGCCAGAGAGTAACCTAGAGCCTTCGGGIGTIGGCTTGAGAGAGGACACCCCA 2440
Db     1435 CRRFFFFRPPRRFRPPRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376

Qy     2441 CAGGAAGCTGAAAGCAAGCCTCAGCGCCCCAAGAGCTTCCAACAGAAACGGGACTATTTC 2500
Db     1375 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316

Qy     2501 CAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACGGTGAAACCTCCAGCCAAGGTTCAG 2560
Db     1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256

Qy     2561 ATCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGATGAGACCGCCGAGTTGTCC 2620
Db     1255 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196

Qy     2621 CCTCCTCCTCCCCCTCCCCCGGTTGTGAAGAAGCCGCTGAAGGCAAGCAGGCCCAAAGCC 2680
Db     1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136

Qy     2681 GTAAAGAAGATGAGGCAGAGCCCGCCCAGGAGGAAGTACCGACCCAGGGCGAGGATCCC 2740
Db     1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076

Qy     2741 CCGGTGCACAGCTCCAACTCCGCACCTCAGCACCCCA 2777
Db     1075 RRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAA 1039
```

RESULT 2
US-08-905-223-162
; Sequence 162, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bacrot, Bruno
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 162:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 56..271
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 3.6
; OTHER INFORMATION: seq CTSLLLYDASNS/EW
US-08-905-223-162

```
  Query Match             1.1%;  Score 46.8;  DB 4;  Length 364;
  Best Local Similarity   56.5%;  Pred. No. 0.039;
  Matches   87;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy      271 TCCCACCCCTGGAAAGATCCCAGGCCTCTTGGACGTGGGGATCTGGTAGAGCAGCTGCC 330
Db       40 TTCAGCTCCAGGCTAGATGCCAGACCAGTTTGACCAGGGGGTTGTGCTGAACCAGCTGCC 99

Qy      331 TCAGGCTGGCATCCTGGAGATCATAGGCACCCGGAGTACCCACTTCCCCCTGCCAGTGTC 390
Db      100 GTACTCAGGGATGCTGGAGACTGTGAGAATCCCCAAACCTCCCTATCCCCTCCCAACACC 159

Qy      391 CTTCCAAGTCTTTCTGGCAAGGTTCCATGCCCTG 424
Db      160 CTTTCAGGACTTTTACAAAAGGTATAAAGTGCTG 193
```

RESULT 3
US-09-535-008-36
; Sequence 36, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H. F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318 259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: US 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 776
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-535-008-36

```
  Query Match             1.0%;  Score 44.8;  DB 4;  Length 776;
  Best Local Similarity   47.0%;  Pred. No. 0.17;
  Matches  139;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy     1075 GCGGTTAGATGGCGAGAACCCTCAGCAGGCTCTGCAGCAT AAGAGAGCAGCGCAGCTT 1134
Db      268 GCAGTGCGCGGGCTTGTCCTCTTCCTCCTACAGCTGGGCCAGGAGCTGGGGTGTCAT 327

Qy     1135 CCTGGGGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACCATGGCAGCTCCCTGGCA 1194
Db      328 GCCGAGGGACACAGGCTTGGAGACAGCCTCAATGCTAAGCCCTACAAGCCCAGCAAGCCG 387

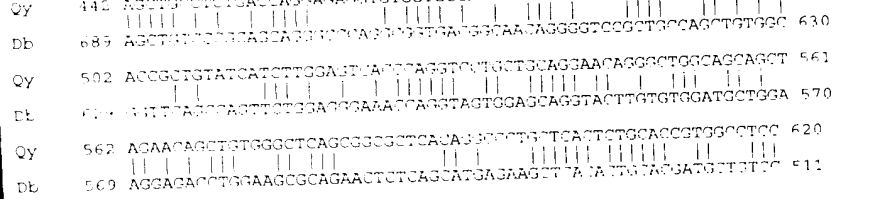
Qy     1195 GCAGGCCCAGCGACGCCCCGGACTCGGACATGAGCTCCAGCTACCTCTGTCTACCT 1254
Db      388 CCAGTCCCTGCGCGAGGCCCGCATCACTGAGAAGCTGGAGAAGACAGAAGATGAGGA 447

Qy     1255 GTGGCGGCAACCCAGATGAGCAACAGAATCAGCGTGGCTGGCGCAAGCGGATGCCCT 1314
Db      448 GGAGCGCAAGCGCCGGCACAAGCACCAGGTACGCTCCCGCGCCCAAGCCCCTGCAGCC 507

Qy     1315 CAGCTCCTTTGCTCCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATCTG 1374
Db      508 CGCCCACCTGGCTGCCTGGCTTGTCCAGGGGTTGCCACGGGGCTGTGTTTCTTTG 563
```

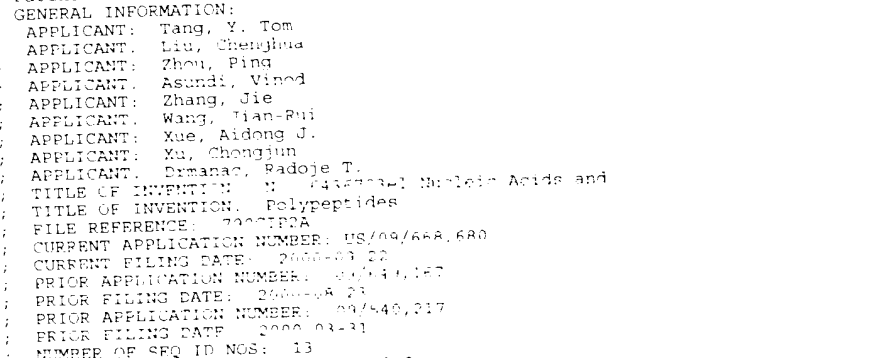
RESULT 4
US-09-668-680-3/c
; Sequence 3, Application US/09668680
; Patent No. 6436703
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenglua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Wang, Tian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Xu, Chongjun
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: N [illegible] Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: [illegible]
; CURRENT APPLICATION NUMBER: US/09/668,680
; CURRENT FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: 2000-[illegible]
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: pt_FI_[illegible] Version 2.0
; SEQ ID NO 3
; LENGTH: 1357
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (284)..(904)
US-09-668-680-3

Query Match 1.0%; Score 43; DB 4; Length 1357;
Best Local Similarity 52.5%; Pred. No. 0.67;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;



RESULT 5
US-09-443-501A-2/c
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,888
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 1.0%; Score 42.4; DB 4; Length 71989;
Best Local Similarity 47.1%; Pred. No. 4.4;
Matches 164; Conservative 0; Mismatches 181; Indels 3; Gaps 1;



RESULT 6
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNEP, F.
; APPLICANT: SCHEIFLINGEP, F
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMPINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDPPSS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22312-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CUPPENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6

```
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-2500
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZapt-Fls
US-08-232-463-14

  Query Match                 1.0%;  Score 41.6;  DB 1;  Length 7218;
  Best Local Similarity   2.1%;  Pred. No. 2.6;
  Matches    8;  Conservative   21;  Mismatches  159;  Indels    0;  Gaps    0;

Qy    3026 TCTCTAGCGGCTGTAGTTCCTCGACCTAAGGCTCCACCTCGTCTTGAGCCTTCCCTATCC 3085
           :::::: ::  :: : ::::::  :::   :::  :::: ::::   :::::::: :::
Db    1167 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1226

Qy    3086 ATCTAGGAAAAGCAGGGACCCCTTCGGGACTTGTTTGGCCCATGTAGTCCAAACCCACCT 3145
           :::   :  :  :::::::   ::: ::: ::: :: :: ::  ::: :::
Db    1227 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1286

Qy    3146 ACAGTTTAGCACCCCCGCTCCAGTAGTACCTCTGTCTGGGGAGCCCAAG 3205
           :  :::: : ::::: :::: ::  : ::::: :: :::: :::      :::
Db    1287 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1346

Qy    3206 ATTCTTTCAGTGGAGTCTCATGCCTTGACAGAGCCCATGGAGGACAAGAACATCTCCACA 3265
           ::::::: :   :::: : :::  :   ::: :      :     : ::::: :
Db    1347 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1406

Qy    3266 AATCTGTTTGTGCCTCTGGAAGTCTCTCTCCATGCCAATGCACCCTGGAAGTTG 3325
           :::::: : ::::::  : : : ::::::: :: : :  ::::   :::::::
Db    1407 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1466

Qy    3326 TTCTTACGCAAGGAGGTGTTCTACCCCGGGAGAACTTCAGTCATCCATACTGCCTCAGT 3385
           ::::: : :       : :::: :::::      :::: :: ::: : :::::  :
Db    1467 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1426

Qy    3386 CTCTCTGCCAGCAGATCCTGC 3407
           ::::::: : ||| || |  |
Db    1427 YYYYYYYYGTACCAAATTCTTC 1448


RESULT 7
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
;    APPLICANT:  Zhang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  Immediate Early Protein From Kaposi's
;    TITLE OF INVENTION:  Sarcoma Associated Herpesvirus, DNA
;    TITLE OF INVENTION:  Encoding Same And Uses Thereof
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036

;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,323A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3489 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3489
US-08-728-323A-1

  Query Match                 2.9%;  Score 41;  DB 2;  Length 3489;
  Best Local Similarity   47.2%;  Pred. No. 2.6;
  Matches  125;  Conservative    0;  Mismatches  140;  Indels    4;  Gaps    3;

Qy    2484 AGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCTGATTAGACTTAAGAAGTTAAAC 2543
           || | | | |  ||   |||| |  ||   |  | | |     | | |  |  |
Db    2021 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATG 2080

Qy    2544 CTCCAGCCAAGGTTCAGATCGCCCAAGAGGAGATCAAGTCAGAAAAGGAGGAGGATG 2603
           |    ||| | ||     || ||||   | |||| | |  | |  |||| |||
Db    2081 AGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATG 2140

Qy    2604 AGACGGCCGAGTTGCCCCTCCTGCTCGCTTCCCCCTTTAAAAAGTTAAC 2653
           ||   | |||  |   |  |  | | | | | | |  |  |  |||| |||
Db    2141 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATG 2200

Qy    2654 CAAGCAGGCCCAAAGCCCTAAAGGAAGATGAGCCAGAGCCCCTAAAAAAGTAGTGA 2711
           |||   | |  |  || |  | |||  |||       | | | |  | | |
Db    2201 AGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCA 2260

Qy    2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
           |||   ||||| |    |  |||
Db    2261 AGCAGCAGGAGGAGCAGGAGCAGCA 2285


RESULT 8
US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT: Ballestas, Mary E.
;  APPLICANT: Kaye, Kenneth M.
;  TITLE OF INVENTION: PHAGING VIRUS LANA ACTS IN TRANS AS A UNIT OF REPLICN
;  TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/298,568
;  CURRENT FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,422
;  EARLIER FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3489
;   TYPE: DNA
```

ORGANISM: Kaposi's sarcoma associated herpesvirus
S-09-298-568-1

Query Match 0.9%; Score 41; DB 4; Length 3489;
Best Local Similarity 47.2%; Pred. No. 6.8;
Matches 125; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 2484 AGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACGGTGAAAC 2543
Db 2021 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 1962
Qy 2544 CTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGATG 2603
Db 1961 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 1902
Qy 2604 AGACCCCCGAGTTGTCCCCTCCTCCTCCCCCTCCCCCGGTTGTGAAGAAGCCGCTGAAGG 2663
Db 1901 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAACAGG 1842
Qy 2664 CAAGCAGGCCCAAAGCCGTAAAGGAAGATGAGGCAGAGCCCGCCCAGGAGGAAGTACCGA 2723
Db 1841 AGCAGCAGCAGCAGGATGAGCAGCAGCAGCAAGATGAGCAGCAGCAGCAGGATGAACAGG 1782
Qy 2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
Db 1781 AGCAGCAGGAGGAGCAGGAGCAGCA 1757

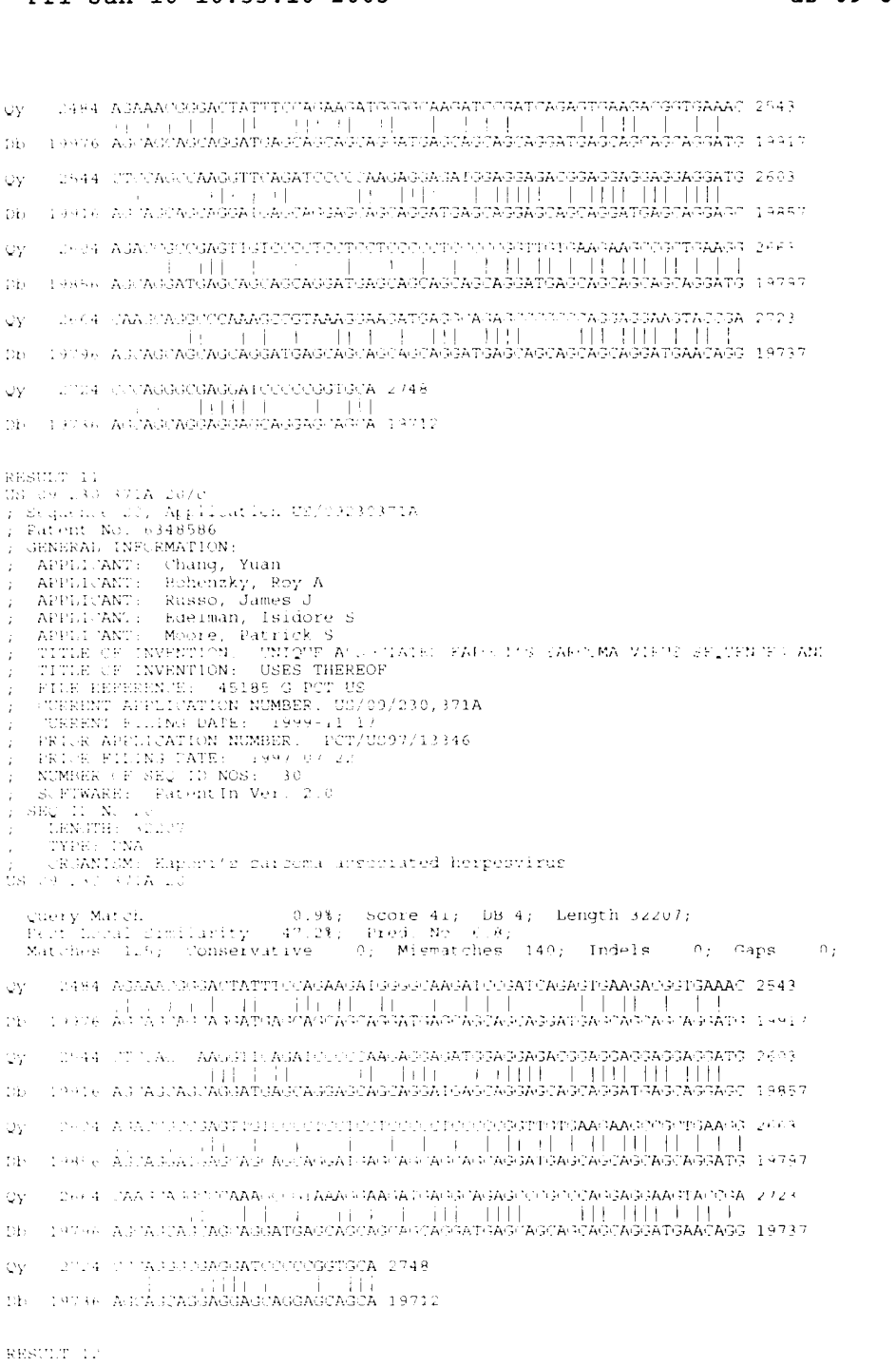
RESULT 9
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52447
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 0.9%; Score 41; DB 2; Length 32207;
Best Local Similarity 47.2%; Pred. No. 6.8;
Matches 125; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 2484 AGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACGGTGAAAC 2543
Db 19976 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 19917
Qy 2544 CTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGATG 2603
Db 19916 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 19857
Qy 2604 AGACCCCCGAGTTGTCCCCTCCTCCTCCCCCTCCCCCGGTTGTGAAGAAGCCGCTGAAGG 2663
Db 19856 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAACAGG 19797
Qy 2664 CAAGCAGGCCCAAAGCCGTAAAGGAAGATGAGGCAGAGCCCGCCCAGGAGGAAGTACCGA 2723
Db 19796 AGCAGCAGCAGCAGGATGAGCAGCAGCAGCAAGATGAGCAGGATGAGCAGCAGCAGGATGAACAGG 19737
Qy 2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
Db 19736 AGCAGCAGGAGGAGCAGGAGCAGCA 19712

RESULT 10
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

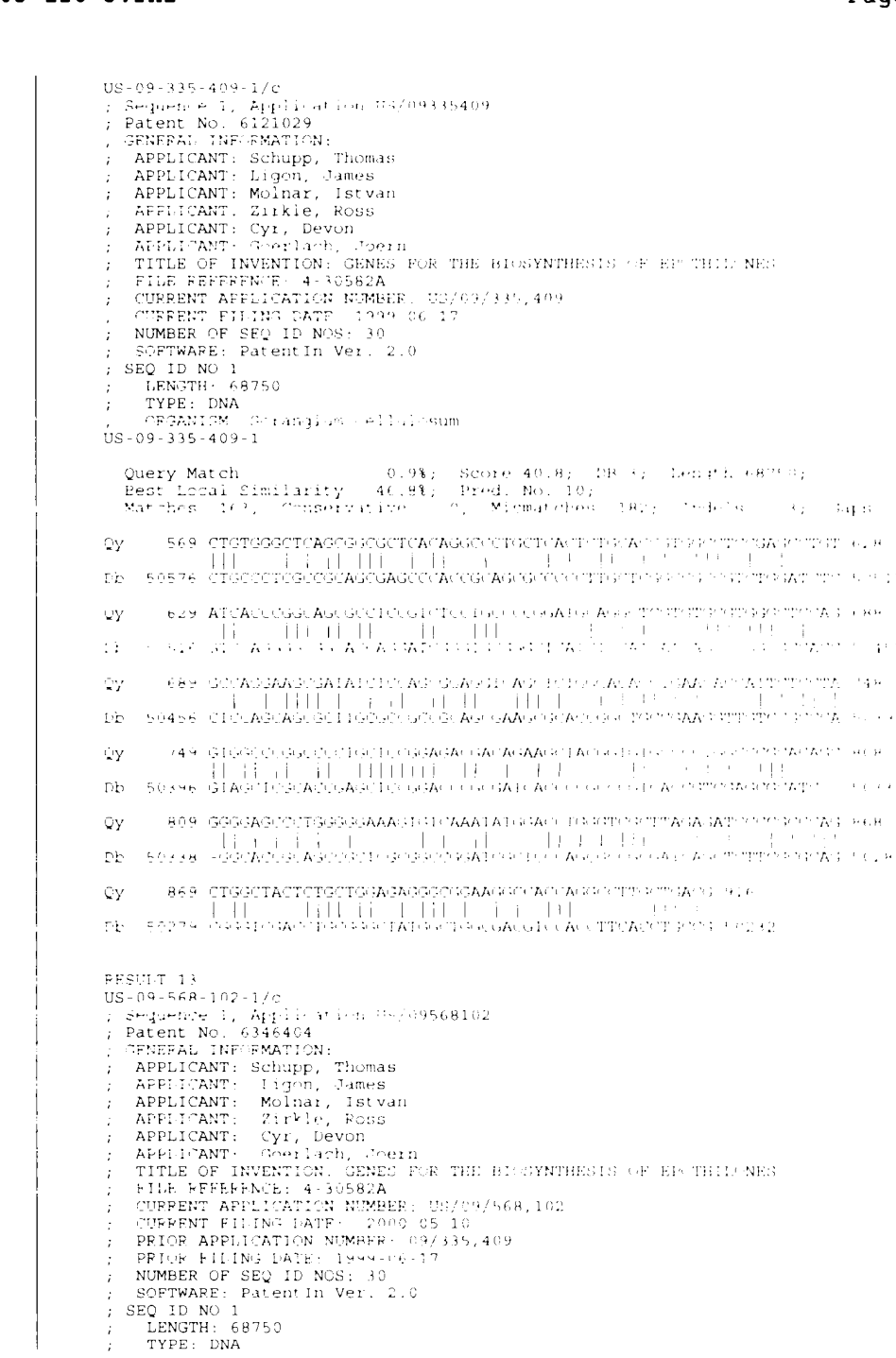
Query Match 0.9%; Score 41; DB 4; Length 32207;
Best Local Similarity 47.2%; Pred. No. 6.8;
Matches 125; Conservative 0; Mismatches 140; Indels 0; Gaps 0;



RESULT 11
US-09-230-371A-26/c
; Sequence 26, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED PAN-HIV VIRUS SEQUENCES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185 G PCT US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-01-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 22207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-26

Query Match 0.9%; Score 41; DB 4; Length 22207;
Pct Local Similarity 47.2%; Prod No 1.8;
Matches 126; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

RESULT 12



RESULT 13
US-09-568-102-1/c
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA

ORGANISM. Sorangium cellulosum
S-09-568-102-1

Query Match 0.9%; Score 40.8; DB 4; Length 68750;
Best Local Similarity 46.8%; Pred No. 10;
Matches 163; Conservative 0; Mismatches 182; Indels 3; Gaps 1;

RESULT 14
US-09-567-969-1/c
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Oberlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 0.9%; Score 40.8; DB 4; Length 68750;
Best Local Similarity 46.8%; Pred No. 10;
Matches 163; Conservative 0; Mismatches 182; Indels 3; Gaps 1;

RESULT 15
US-09-568-480-1/c
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION.
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Oberlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 0.9%; Score 40.8; DB 4; Length 68750;
Best Local Similarity 46.8%; Pred. No. 10;
Matches 163; Conservative 0; Mismatches 182; Indels 3; Gaps 1;

Search completed: January 10, 2003, 09:29:28
Job time : 1109.28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:18:37 ; Search time 168.45 Seconds
(without alignments)
11430.411 Million cell updates/sec

Title: US-09-803-126-3
Perfect score: 4375
Sequence: 1 cgctgggactgtcacctacc..... aaaaaaaaaaaaaaaaaaaa 4375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38[illegible] seqs, 2[illegible] residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4375	100.0	4375	10	US-09-803-126-3	Sequence 3, Appli
2	4352	99.5	6293	10	US-09-803-126-2	Sequence 2, Appli
3	946.4	21.6	3780	10	US-09-803-126-7	Sequence 7, Appli
4	946.4	21.6	4134	10	US-09-803-126-5	Sequence 5, Appli
5	621.2	14.2	1247	9	US-09-764-868-145	Sequence 145, App
6	166.2	3.8	24757	9	US-09-764-868-1347	Sequence 1347, Ap
C 7	89.4	2.0	656	10	US-09-864-761-7087	Sequence 7087, Ap
C 8	75.2	1.7	170	10	US-09-864-761-23818	Sequence 23818, A
9	49.4	1.1	3856	12	US-10-002-600-7	Sequence 7, Appli
10	49.4	1.1	7465	10	US-09-880-107-3357	Sequence 3357, Ap
11	47.4	1.1	145	10	US-09-882-540-3077	Sequence 3077, Ap
12	47	1.1	316	10	US-09-864-761-17473	Sequence 17473, A
13	47	1.1	436	10	US-09-864-761-686	Sequence 686, App
14	44	1.0	5520	9	US-09-902-525-38	Sequence 38, Appl
15	43.2	1.0	2038	10	US-09-880-107-2102	Sequence 2102, Ap
16	43	1.0	2844	9	US-09-905-291A-48	Sequence 48, Appl
17	43	1.0	2822	9	US-09-902-853-48	Sequence 48, Appl
18	43	1.0	2822	9	US-09-907-824-48	Sequence 48, Appl
19	43	1.0	2822	9	US-09-907-841-48	Sequence 48, Appl
20	43	1.0	2822	9	US-09-904-011-48	Sequence 48, Appl
21	43	1.0	2822	9	US-10-028-072-305	Sequence 305, App
22	43	1.0	2822	10	US-09-828-366-6	Sequence 6, Appli
23	43	1.0	2822	10	US-09-909-320-48	Sequence 48, Appl
24	43	1.0	2822	10	US-09-909-088B-48	Sequence 48, Appl
C 25	42.8	1.0	2379	10	US-09-781-080B-12	Sequence 12, Appl
26	42.8	1.0	2802	10	US-09-925-300-238	Sequence 238, App
27	41.8	1.0	162	9	US-10-046-935-933	Sequence 933, App
28	41.8	1.0	162	9	US-09-878-178-933	Sequence 933, App
C 29	41.8	1.0	461	10	US-09-864-761-14656	Sequence 14656, A
C 30	41.8	1.0	526	10	US-09-864-761-31189	Sequence 31189, A
31	41.4	0.9	1960	10	US-09-822-849A-265	Sequence 265, App
32	41.4	0.9	4509	9	US-10-124-800-5	Sequence 5, Appli
33	41.4	0.9	4512	9	US-10-124-800-27	Sequence 27, Appl
34	41	0.9	440	10	US-09-864-761-3517	Sequence 3517, Ap
C 35	40.8	0.9	68750	9	US-10-014-717-1	Sequence 1, Appli
36	40.6	0.9	1267	10	US-09-739-907-38	Sequence 38, Appl
37	40.6	0.9	6604	10	US-09-880-107-1748	Sequence 1748, Ap
C 38	40.2	0.9	4080	10	US-09-778-927A-16	Sequence 16, Appl
C 39	40.2	0.9	4366	12	US-10-044-090-868	Sequence 868, App
C 40	40.2	0.9	659158	9	US-09-771-208-20	Sequence 20, Appl
41	40	0.9	349	10	US-09-864-761-4616	Sequence 4616, Ap
42	40	0.9	2074	10	US-09-879-957-19	Sequence 19, Appl
43	39.8	0.9	177	10	US-09-864-761-20287	Sequence 20287, A
44	39.6	0.9	956	10	US-09-925-301-282	Sequence 282, App
45	39.6	0.9	3240	10	US-09-119-855-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-803-126-3
; Sequence 3, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT. Deng, Gary G.
; APPLICANT Kubanyi, Gabor M
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4375
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse myosin related protein (MRP) variant 2 cDNA
US-09-803-126-3

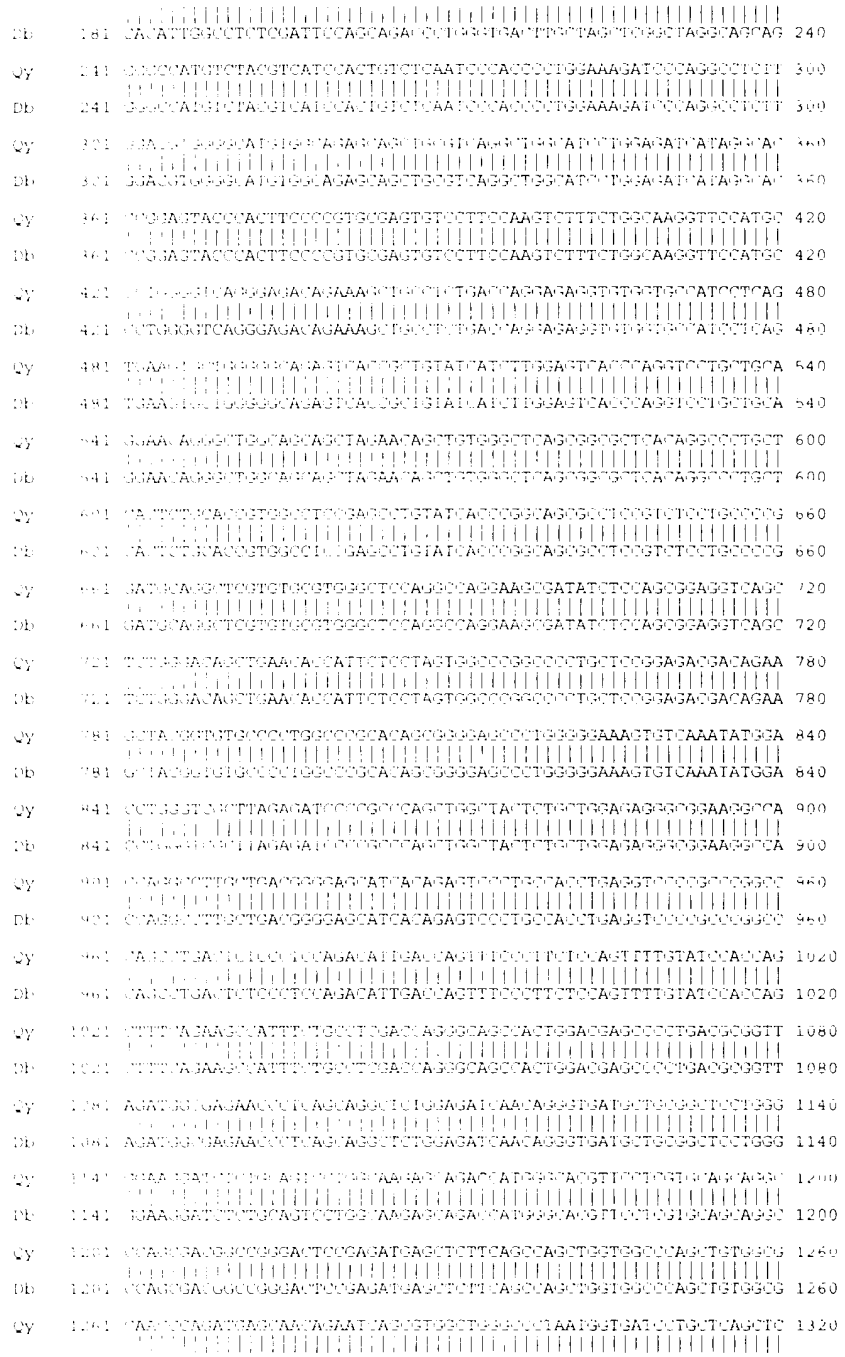
Query Match 100.0%; Score 4375; DB 10; Length 4375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

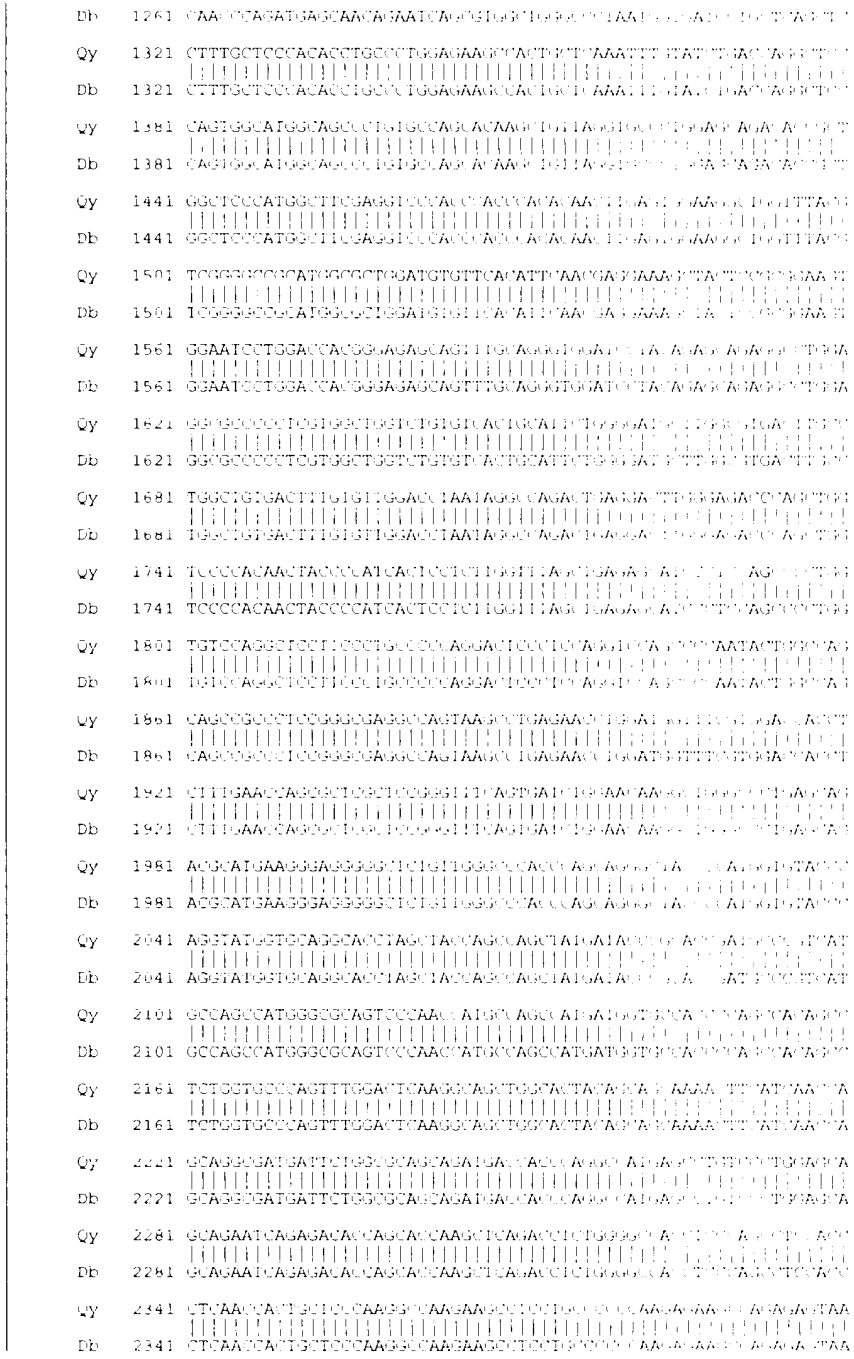
Qy 1 CGCTGGGACTGTCACCTACCAGGTGCACAAGTTCATAAACAGAAACAGGGGCCACCTGGA 60
Db 1 CGCTGGGACTGTCACCTACCAGGTGCACAAGTTCATAAACAGAAACAGGGGCCACCTGGA 60

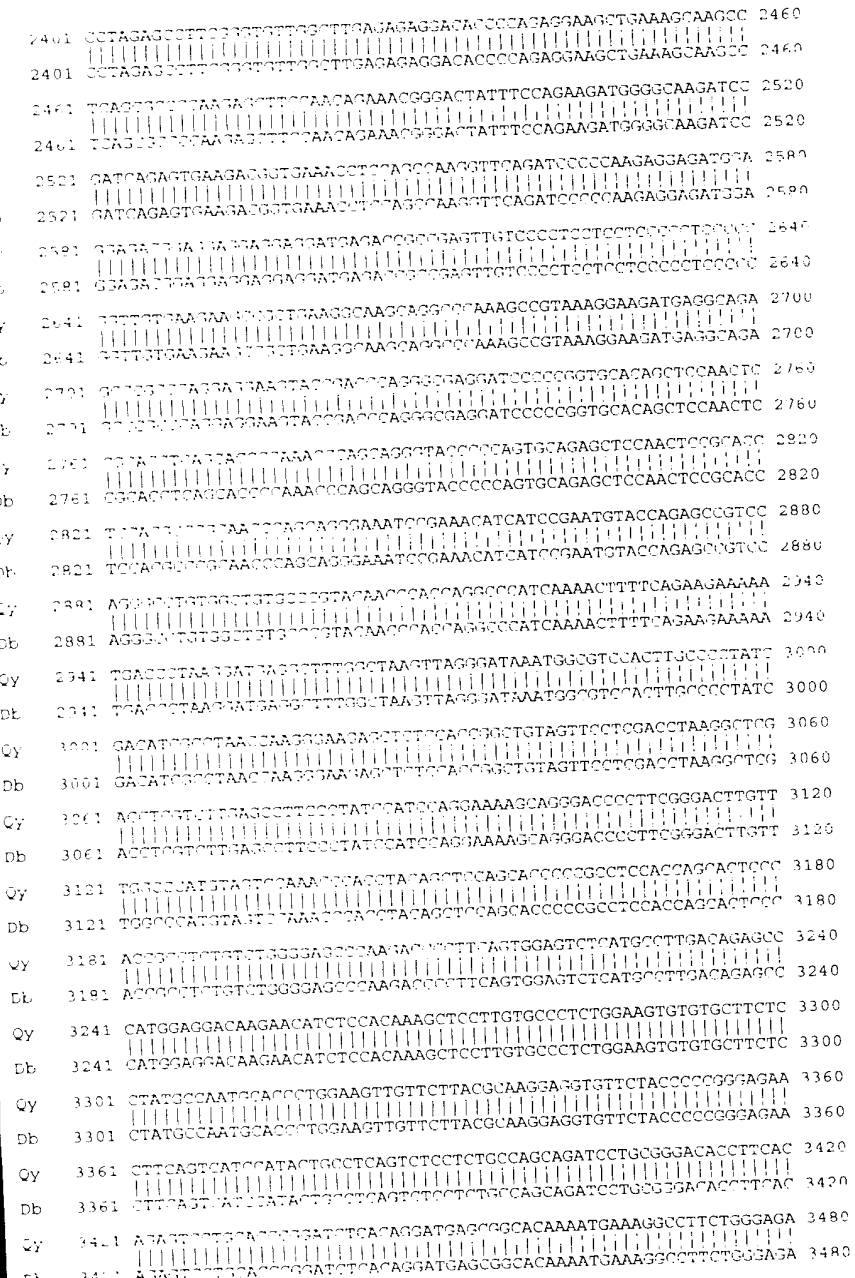
Qy 61 CCCCGCTGTGCTGGAGATGCTCAGGCAGAGCCAGCTGCAGGTGACCTAGCCTTCCTTTCA 120
Db 61 CCCCGCTGTGCTGGAGATGCTCAGGCAGAGCCAGCTGCAGGTGACCTAGCCTTCCTTTCA 120

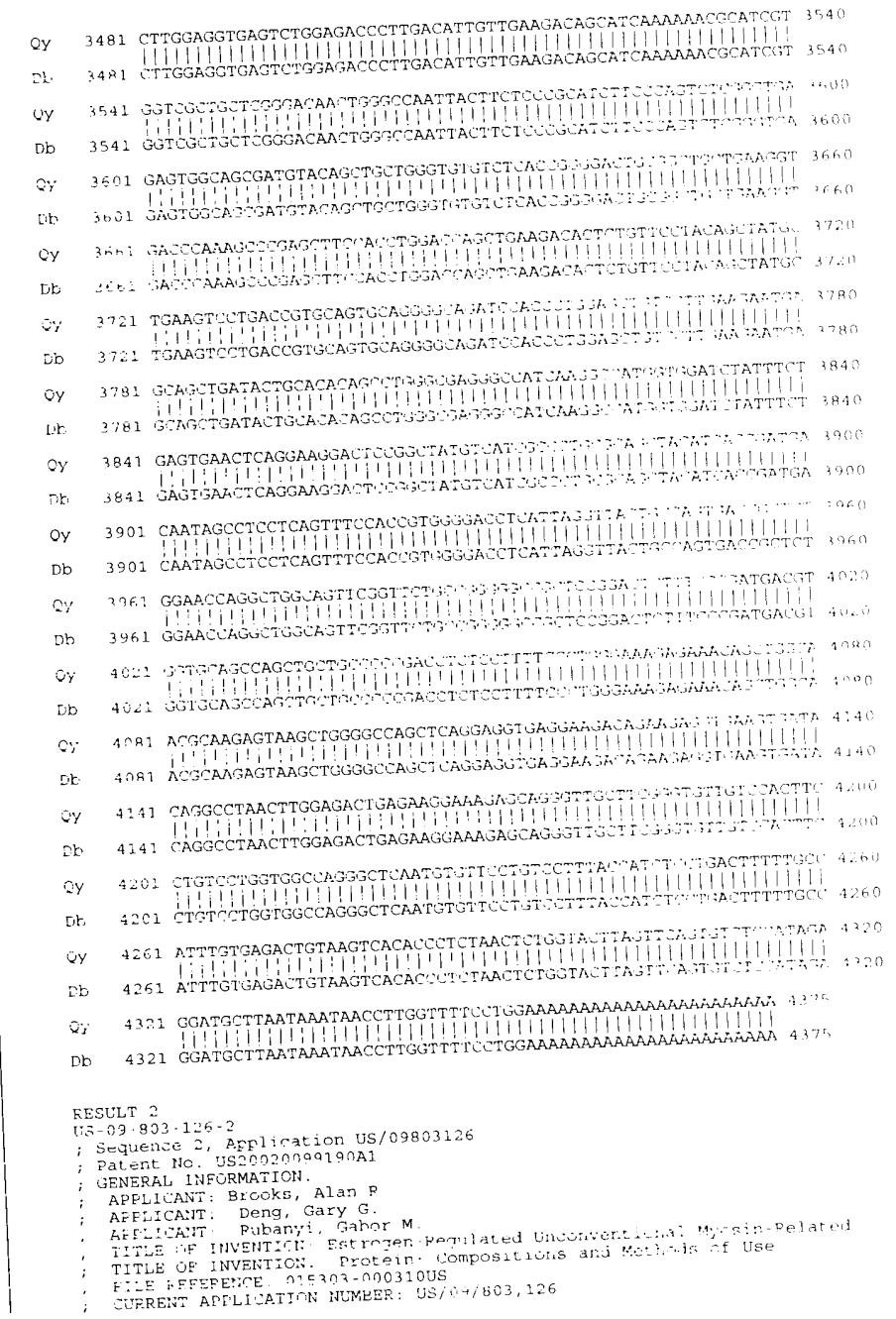
Qy 121 GCTCATGGGCAGCCTGTTCCAAGAAGCAGAGCCCCAGGCTGGGACTGAGCAAAACAAACC 180
Db 121 GCTCATGGGCAGCCTGTTCCAAGAAGCAGAGCCCCAGGCTGGGACTGAGCAAAACAAACC 180

Qy 181 CACATTGGCCTCTCGATTCCAGCAGAGCCTGGGTGACTTCTTAGCTCAGGTAGGCAGCAG 240





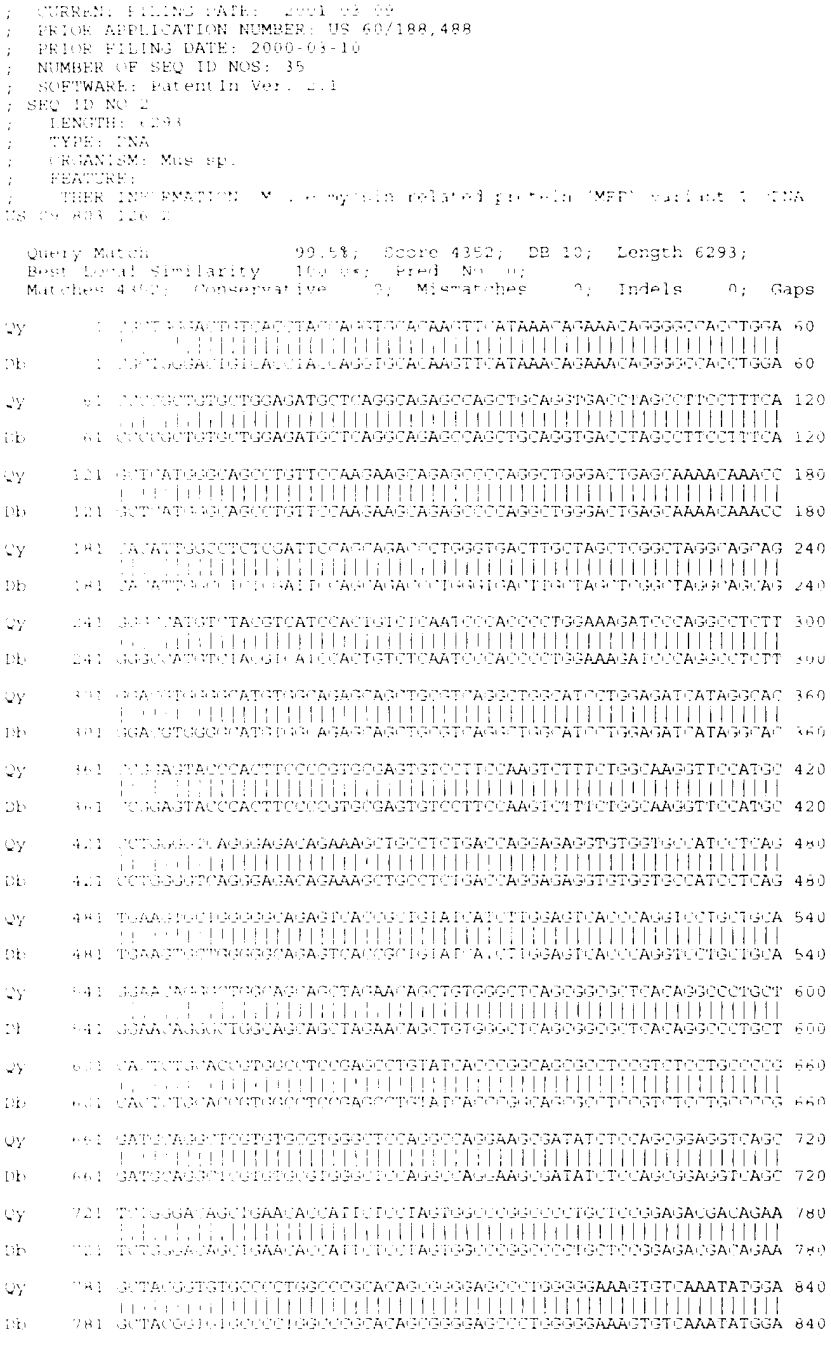


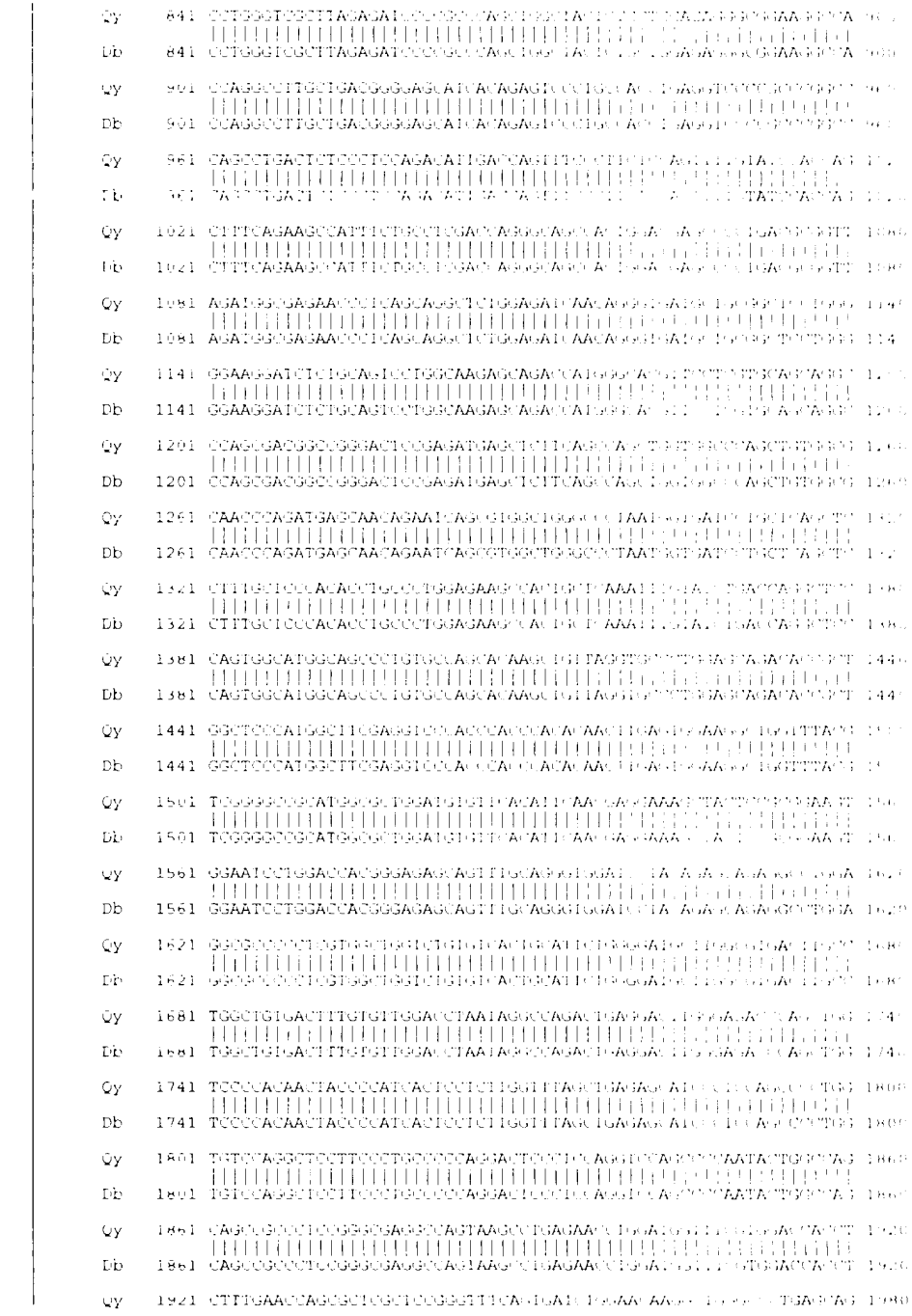


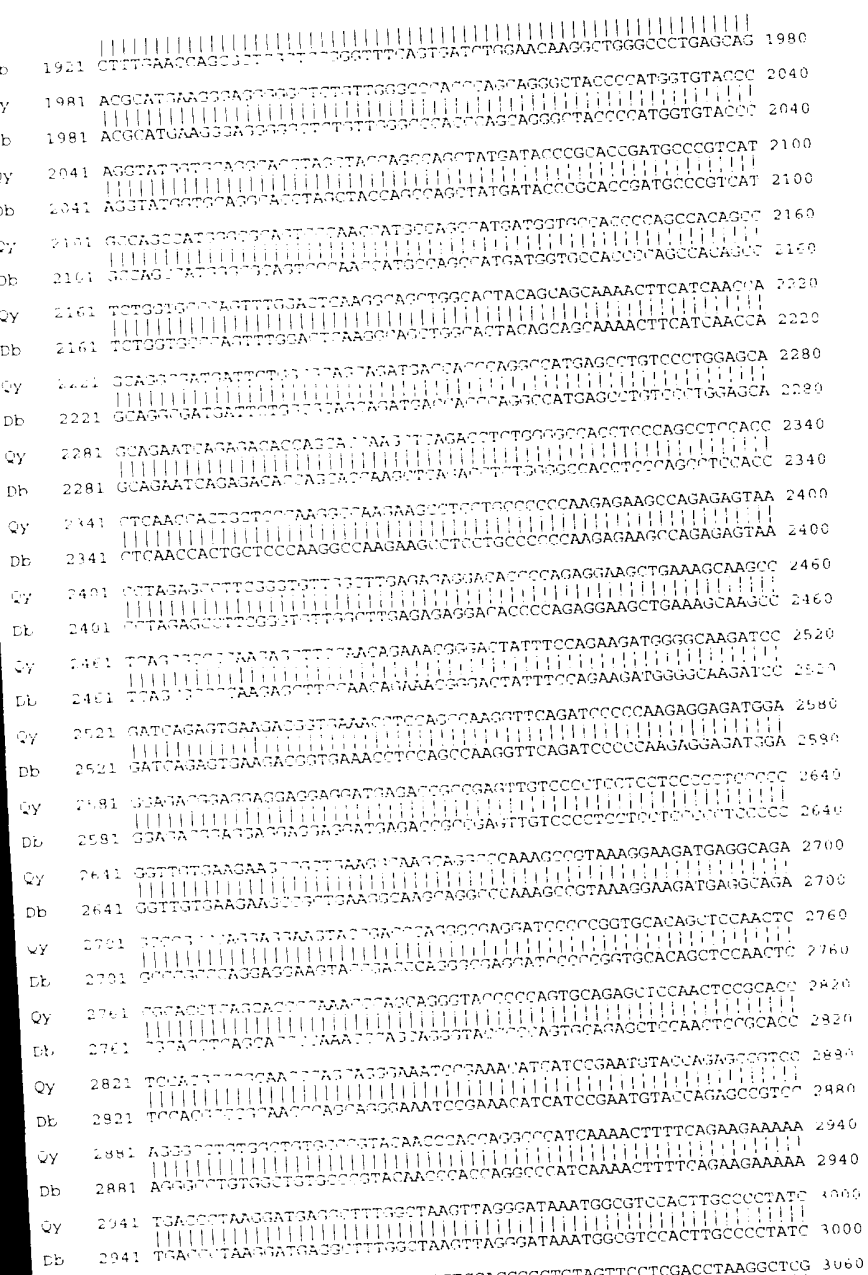
RESULT 2
US-09-803-126-2
; Sequence 2, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION.
; APPLICANT: Brooks, Alan R
; APPLICANT: Deng, Gary G.
; APPLICANT: Pubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126

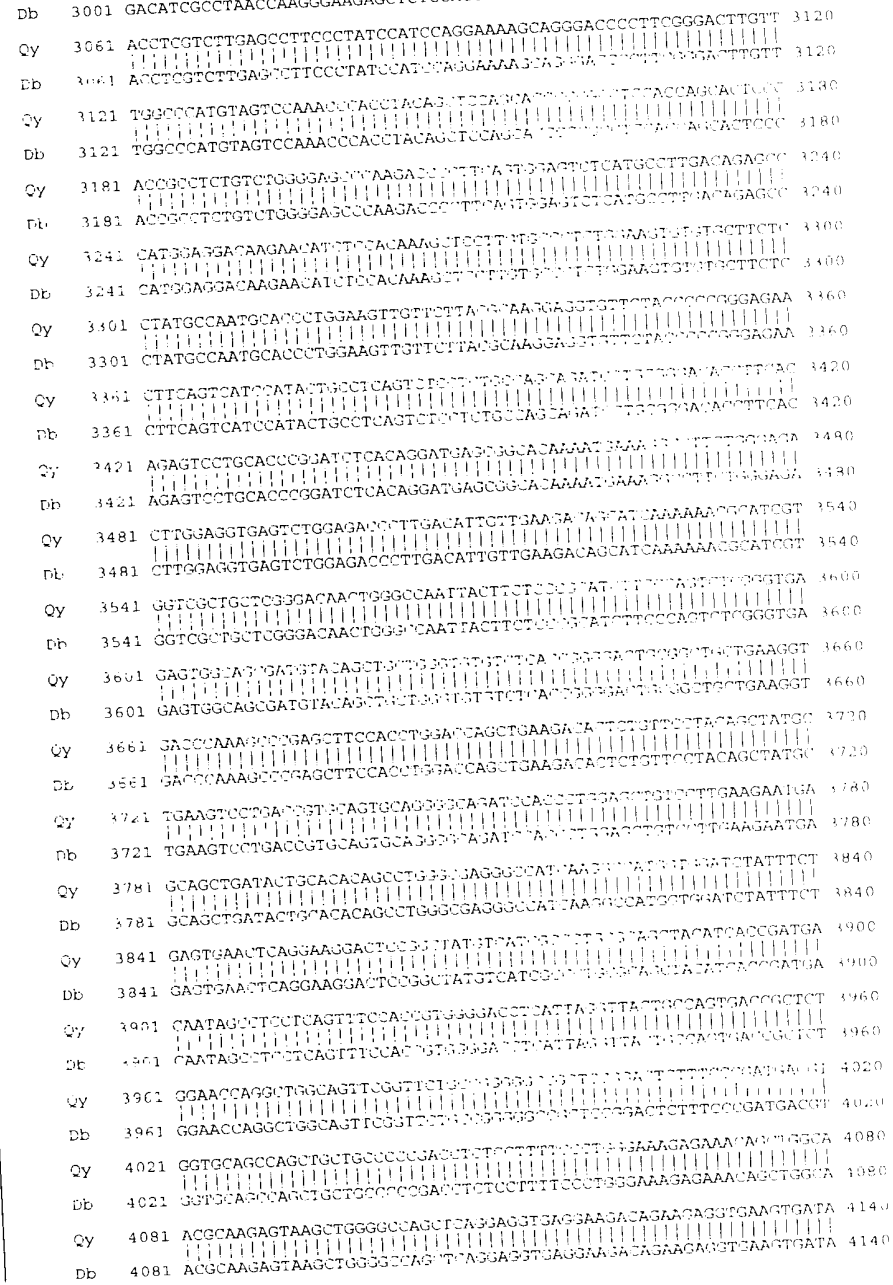
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 6293
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Myosin related protein 'MRP' variant 1 cDNA
US-09-803-126-3

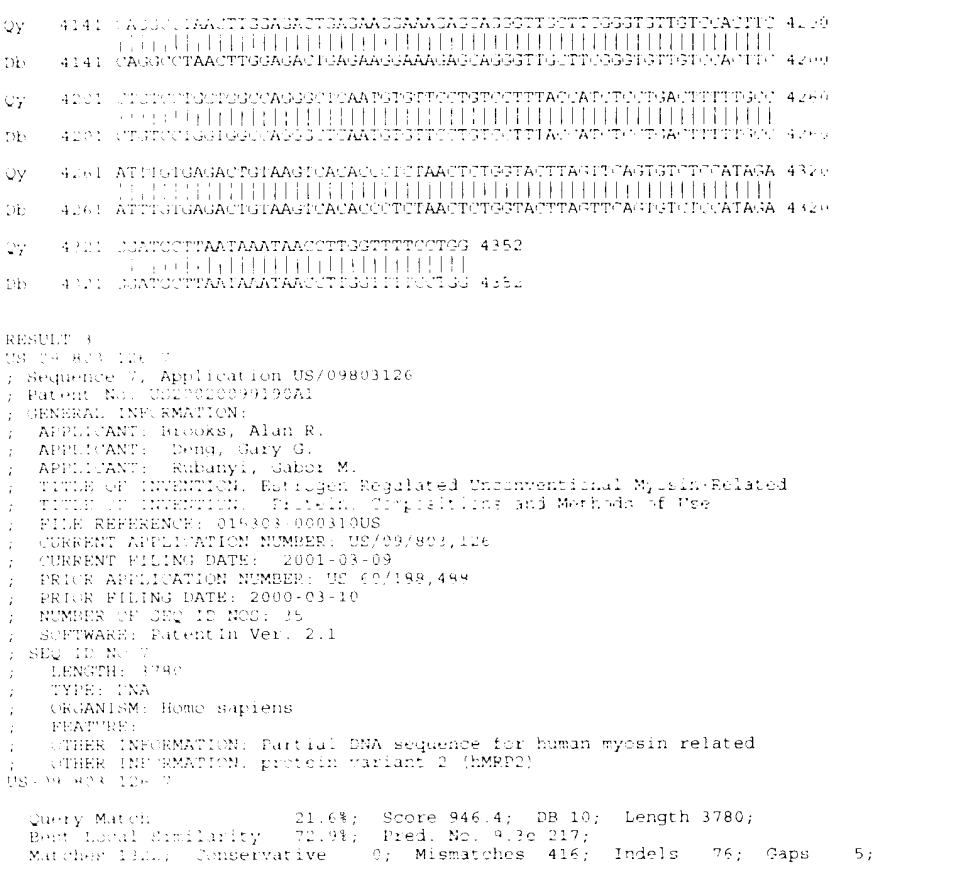
Query Match 99.6%; Score 4352; DB 10; Length 6293;
Best Local Similarity 100.0%; Pred No. 0;
Matches 4352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





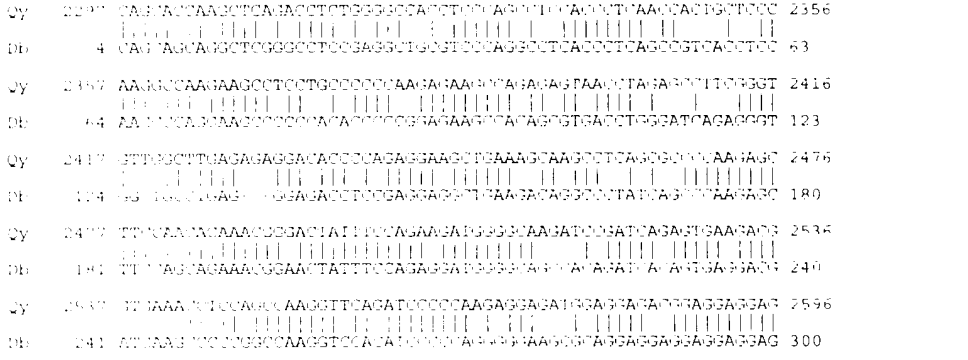


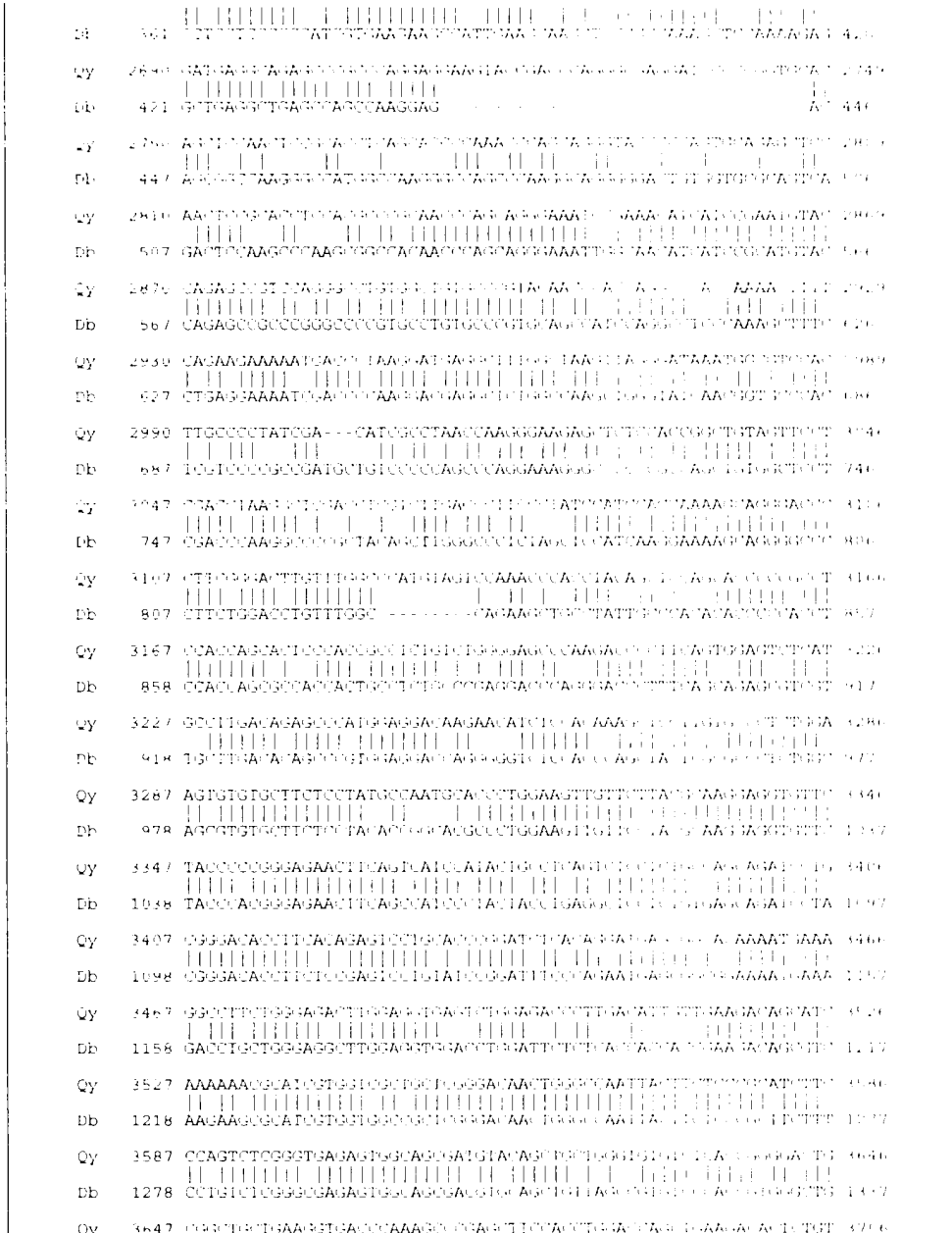


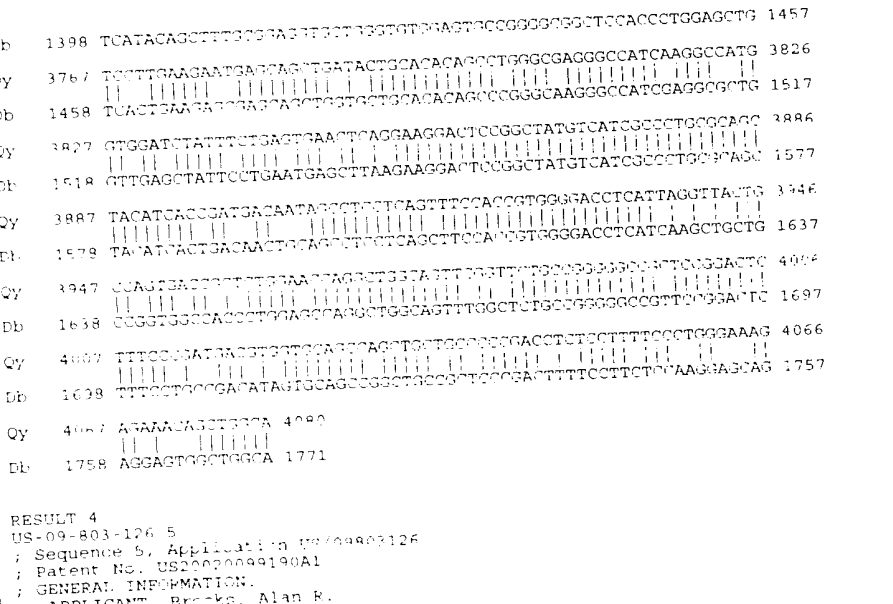


RESULT 3
US-09-803-126-7
; Sequence 7, Application US/09803126
; Patent No. US20020090190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Dona, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015308-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related
; OTHER INFORMATION: protein variant 2 (EMRP2)
US-09-803-126-7

Query Match 21.6%; Score 946.4; DB 10; Length 3780;
Best Local Similarity 72.9%; Pred. No. 9.3e-217;
Matches 1122; Conservative 0; Mismatches 416; Indels 76; Gaps 5;

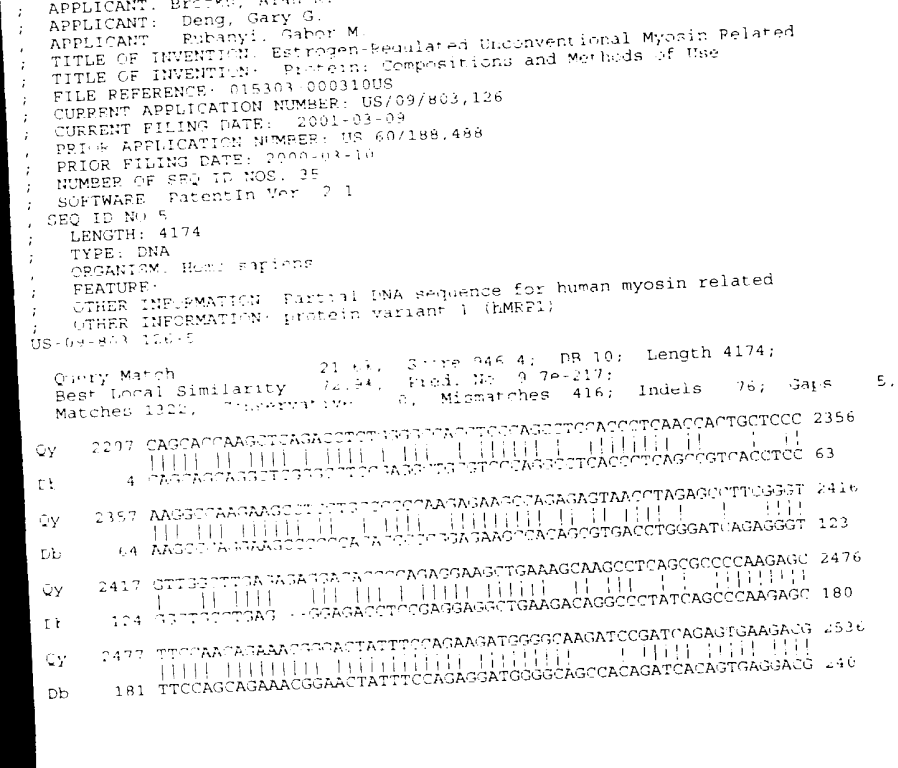






RESULT 4
US-09-803-126-5
; Sequence 5, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related
; OTHER INFORMATION: protein variant 1 (hMRP1)
US-09-803-126-5

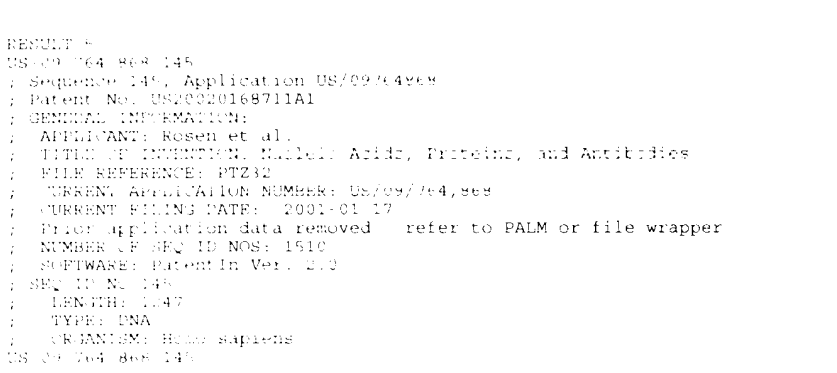
Query Match 21.4%; Score 946.4; DB 10; Length 4174;
Best Local Similarity 74.9%; Pred. No. 9.7e-217;
Matches 1322; Conservative 0; Mismatches 416; Indels 76; Gaps 5.

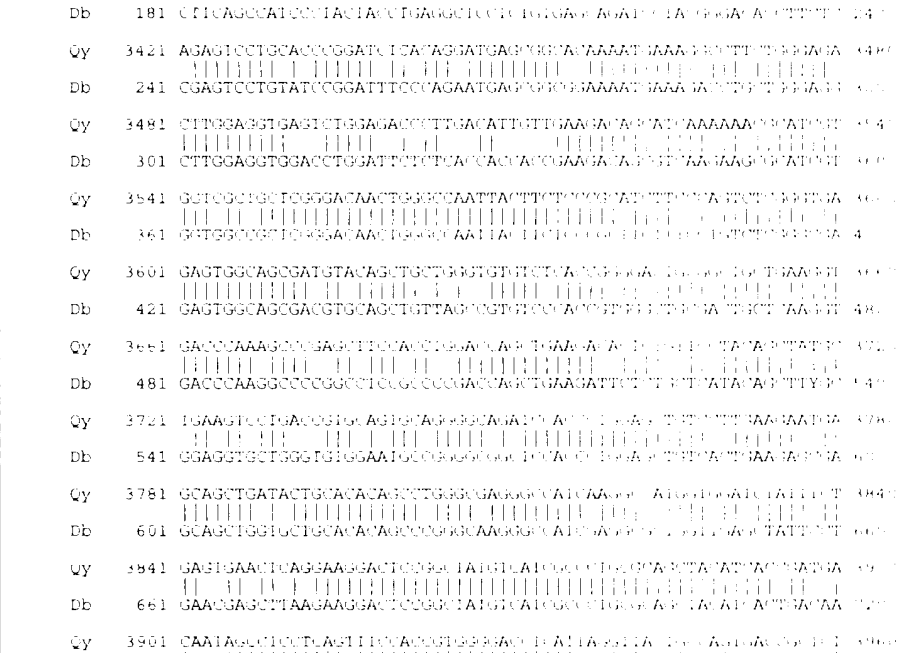




RESULT 5
US-09-764-868-145
; Sequence 145, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 145
; LENGTH: 1647
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-145

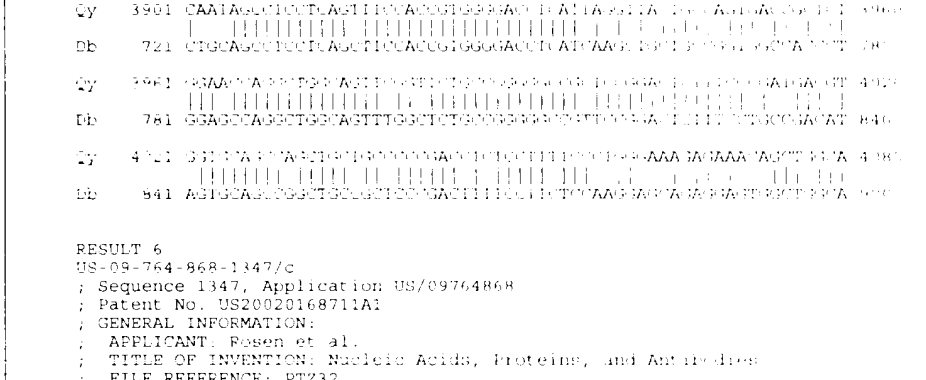
Query Match 14.1%; Score 1247; DB 9; Length 1647;
Best Local Similarity 80.6%; Pred. No. 3.9e-139;
Matches 725; Conservative 1; Mismatches 174; Indels 0; Gaps 0;





RESULT 6
US-09-764-868-1347/c
; Sequence 1347, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1347
; LENGTH: 24757
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1347

Query Match 3.8%; Score 166.2; DB 9; Length 24757;
Best Local Similarity 80.4%; Pred. No. 1.3e-29;
Matches 195; Conservative 0; Mismatches 48; Indels 0; Gaps 0;



```
y   1179 GCAGTTCCTCTTCTTAGTAGCCCCAGCCAGCCCCCCGACTCCGAGATGAGCTCTTCAGCC 1238
         || | | ||| |||| |||||  ||| | |||||  || || ||  |||||||| |||
o  17592 GCGCATAGCTGCTGTAGTGAGCAGCCCAGTGGCCCCCAGGGCTGCGGAATGAGCTCTTTAGCC 17533

y   1239 AGCTCGATTTTTAGCTTCTCCCCTAACCCAGATCAGCAACAGAATCAGCGTCTCTGGGCCC 1298
         ||||||||||||||||| ||||  |||||||||||| || | |||  |||||||||||||
b  17532 AGCTGGTGGCCAAGTATGGCAGAAGCCAGATGAACAGCAGAGCCAGCGTGCTGGGCCC 17473

y   1299 TAATGGTGATCCTGCTCAGGTCCTTTCTCCCACAGCCTGCCCTGGAGAAGCCACTGCTCA 1358
         | ||||   | |||||||| |||||  |||  ||| |||  |||||||||||||||
b  17472 TCATCCTGTTTTGGTCAGCCCTTTCCCCACTGCCTGTCCTACAGAAGCCACTGCTCA 17413

y   1359 AAT 1361
         |||
b  17412 AGT 17410


RESULT 7
US-09-864-761-7087/c
; Sequence 7087, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 40317
;  SOFTWARE: Aeomax Sequence Listing Engine Vers  1.1
; SEQ ID NO 7087
;   LENGTH: 526
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION:  EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION:  EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;   OTHER INFORMATION:  EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION:  EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;   OTHER INFORMATION:  EXPRESSED IN LUNG, SIGNAL =  2
;   OTHER INFORMATION:  EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;   OTHER INFORMATION:  EXPRESSED IN BRAIN, SIGNAL = 1.5
US-09-864-761-7087

  Query Match             2.0%;  Score 89.4;  DB 10;  Length 526;
  Best Local Similarity   86.1%;  Pred. No. 5.1e-12;
  Matches   99;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy   3855 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACCGATGACAATAGCCTCCTCA 3914
          ||||||||||||||||||||||||||||||||||||||||||||||||||  ||  ||||||||||
Db    415 AGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAACTCCAGCCTCCTCA 356

Qy   3915 GTTTCCACCGTGGGGACCTCATTAGCTTACTGCCAGTGACCGCTCTGGAACCAGG 3969
          | ||||||||||||||||||| | | | |||| ||| || | ||||| |||||
Db    355 GCTTCCACCGTGGGGACCTCATCAAGCTTTCTCCCGTGACTCACCCTCCAACCCAGG 301


RESULT 8
US-09-864-761-23818/c
; Sequence 23818, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
```

```
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23818
;   LENGTH: 170
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;   OTHER INFORMATION: EST HUMAN HIT, AW474878.1, EVALUE 1.00e-91
;   OTHER INFORMATION: SWISSPROT HIT, P05496, EVALUE 1.00e-88
;   OTHER INFORMATION, NT HIT, g11426247, EVALUE 1.50e-80
US-09-864-761-23818

  Query Match              1.2%;  Score 75.2;  DB 10;  Length 170;
  Best Local Similarity   79.8%;  Pred. No. 7.6e-09;
  Matches   99;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy     4029 CATGCAGTTCGGTTCTGCCGGGGCCGCTCCGGACTCTTTCCCATGACCTGTGTCAGC 4088
           ||| |||| || ||||||||||||| ||||||||||||| | ||| | |||||||
Db      120 GCTGGCAGTTGGCTCTGCCGGGGCGGTCCGGACTCTTTGCCATGACATAGTGCAGC 177

Qy     4089 TCCTCTTGCCCCCGACCTCTCCTTTTCCTCTGGAAAGAGAAACAGCTGGCA 4080
            | ||||  || |||||| | |||||  |||  ||   |||| |   |||||||
Db      178 GGTCTCCGGCTCCCGACTTTCCTCTCCCAAGGAGCAGAGGAGTGGCTGGCA 59


RESULT 9
US-10-002-603-7
; Sequence 7, Application US/10002603
; Patent No. US20020137077A1
; GENERAL INFORMATION:
;  APPLICANT: Hopkins, Christopher M.
;  APPLICANT:  Peterson, David P.
;  APPLICANT:  Cocks, Benjamin G.
;  APPLICANT:  Hawkins, Phillip R.
;  TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
;  FILE REFERENCE: PA-0042 US
;  CURRENT APPLICATION NUMBER: US/10/002,603
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 60/243,521
;  PRIOR FILING DATE: 2000-10-25
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PERL Program
; SEQ ID NO 7
;   LENGTH: 3836
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Template ID: 1006159.4
US-10-002-603-7

  Query Match              1.1%;  Score 49.4;  DB 12;  Length 3836;
  Best Local Similarity   51.4%;  Pred. No. 0.048;
  Matches  111;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     1119 TCTCTGCATCGTTGGCAAGAGCAGACCATGGGCACGTTCCTCGTGCAGCAGAGCCAGCAA 1207
            || | || | ||||| |  | |  |  |  |  || |  | |||||| ||
Db     2145 TTTGCAGGCTGGGCACATCTGAACTCAGATCTTCACACTGGCCCTGCAG 2204

Qy     1208 CAGCAGGACTCCGAGATGAGCTTCTTGAGCCAGCTGGTGGCCCAGCTGTGGCGCAACCCA 1267
            |||  | ||| || || ||| | |||| | ||  |||||  || ||||| |
Db     2205 CACTGTGCTCTTCCAGGACCAGTTTTACTTCTACTATCCTGAATCAGTTGACAACTCC 2264

Qy     1268 GATGAGCAACAGAATCAGCGTGCTGGCCCTAAATCTCTCAGTCCTTTCCCT 1317
            |  |||     |  |||| ||||||  ||  ||  | |          ||  |||  |
Db     2265 AACAGCCACAGCAAGAGACGGGCTCGCAGCTGCTGTCGCTCTCCACGCCCTTTCCCT 2324

Qy     1328 CCCACACCTGCCCTGGAGAAGCCACTCCTCAAATTTCCTA 1367
            |||      |  |||  |   |      |  || ||| ||
Db     2325 CCCAGCAAGGGGCTGCTGCCCCATGCCCAGAAGTTTATA 2363


RESULT 10
US-09-880-107-3357
; Sequence 3357, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vinkley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,879
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3357
;   LENGTH: 7465
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. [illegible]
US-09-880-107-3357

  Query Match              1.1%;  Score 49.4;  DB 1;  Length 7465;
  Best Local Similarity   51.8%;  Pred. No. 0.066;
  Matches  141;  Conservative    0;  Mismatches  131;  Indels    7;  Gaps    1;

Qy      158 GCTGGGACTGAGCAAAACAAACCCACATTGGCCTCTCAATTCCACATACACCCTCCTTCACT 217
            || | |||   | |   |  |||||  |  |  |  |  | ||  |  |  |||  |||
Db     2073 GCCGAGACCAGGAAGCGCTCGCCACACTTTACACACCTACTTAAGGTCACTCGACTTTCCT 2014

Qy      218 TTCCTACCTCCGCTACACCACCAGCCCCATCCTCACCTCATCCACCTCCCAATCCCCACT 277
            || |      ||| ||   |    ||||   || | |  |  |  |  |||  |||
Db     2133 CTGATGGGTACGTTGGGTGGT   CCCACCCTTTTCTCCCTCTCACCAACCCCAAT 2190

Qy      278 CCTGGAAAGATCCCAGGCCTCTTGACGTGGGCACTCTGCCACAGCACACTCCTCCACC 337
            ||||  ||    || ||| ||  |  |  |  |  |  |  |  |  |  |  |
Db     2191 GAGTTCAAGAAGCCCATCTCTTCGACCGCACCTCCTGAGTCGCCACACTCCACACTCCACC 2250

Qy      338 CGCATCCTGGAGATCATAGGCACCCGGAGTACCCACTTCCCCCCCGCACTCCCCTTCCAA 397
            || ||  |||||| |||  | | |||  |  |  || |||  |  |  |  |||  |
Db     2251 CGAATGATGAGACCATCGAATCCGCCAGTGGCTACCCCACCCCACCAGCTCCCCA 2309

Qy      398 GTCTTTCTGGCAAGGTTCCATGCCCTCGGGTCAGG 432
            |  |||  |||   ||| || ||  |||  | ||||
Db     2310 GAGTTTCCCAGCCCTATGCCCCAGCCAGG 2344


RESULT 11
US-09-783-590-3077
; Sequence 3077, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Baseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Secreted Proteins and Polynucleotides Enc
;  FILE REFERENCE: PC0-16.2C1
```

; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: US/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3077
; LENGTH: 145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (29)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (57)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (138)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (139)
; OTHER INFORMATION: n equals a,t,g, or c
;
US-09-783-590-3077

Query Match 1.1%; Score 47.4; DB 10; Length 145;
Best Local Similarity 54.4%; Pred. No. 0.032;
Matches 93; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 3990 GGGGGGGTTCCGGACTCTTTTCCCGATGACGTGGTGTAGCCAGCTGCTGCGTCTGGACCTTT 4049

Db 15 GGGGGCCGTTCCGTNTCTTTTCTGCCGACATAGTGCAGCAGGNTGCGGCTGCAGATTTT 74

Qy 4050 CCTTTTCCCTGGGAAAGAGAAACAGCTGGCA 4080

Db 75 CCTTCTCCAAGGAGCAGAGGAGTCGCTGGCA 105

RESULT 12
US-09-864-761-17473
; Sequence 17473, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,959
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17473
; LENGTH: 316
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007050.23
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 33
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 39
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 21
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
; OTHER INFORMATION: EXPRESSED IN U87G, SIGNAL = 24
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 24
; OTHER INFORMATION: EST_HUMAN HIT: gi|1478[illegible], EVALUE 1.00e-04
; OTHER INFORMATION: NT HIT: gi|1482953, EVALUE 9.00e-10
; OTHER INFORMATION:
US-09-864-761-17473

Query Match 1.1%; Score 47; DB 10; Length 316;
Best Local Similarity 54.3%; Pred. No. 0.057;
Matches 95; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 2137 GATGGTGCCACCCCAGCCACAGCCTCTGGTGCCCAGTTTGGACTCAAGGCAGCTGGCACT 2196

Db 20 GGTGGCGCTGCAGCAGCAGCAGCAACAGCAGCAGTTCCAGCAGCAGCAGCAGCCGCGCT 79

Qy 2197 ACAGCAGCAAAACTTCATCAACCAGCAGGCGATGATTCTGGCGCAGCAGATGACCACCCA 2256

Db 80 ACAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGTTCCAGCAGCAGAGTGCCATGCA 139

Qy 2257 GCCCATGAGCCCTCTCCTGGAGCAACAGAATCAGACACAGCAGCACCAAGCTCAG 2311

Db 140 GCAGCAGTTCCAAGCAGTAGTCCAGCAGCAGCAGATT[illegible]AGCAGCAG 194

RESULT 13
US-09-864-761-686
; Sequence 686, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 60/235,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,353
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,409
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,113
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 696
;    LENGTH: 436
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO AC007050.23
;    OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  25
;    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  18
;    OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.4
;    OTHER INFORMATION: EXPRESSED IN B.DS MAST.W, SIGNAL =  23
;    OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  39
;    OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  12
;    OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  21
;    OTHER INFORMATION: EXPRESSED IN PT474, SIGNAL =  24
;    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  24
;    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  24
US-09-864-761-696

  Query Match              1.1%;  Score 47;  DB 10;  Length 436;
  Best Local Similarity   94.3%;  Pred. No. 0.055;
  Matches   99;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    2137 GATGGTGGTGCCCCAGCCACAGCCTCTGGTGCCCAGTTTGGACTCAAGGCAGCTGGCACT 2196
           |  |||| |  |||| |||| | | ||        | |      ||||  ||| ||
Db     256 GGTGCCGCTGCAGCAGCAGCAACAGCAGCAGTTCCAGTACCAGCAGCAGGTGGCGCT 315

Qy    2197 ACAATGGCAAAAGTTCATCAACCAGCAGGCGATGATTCTGGCGCAGCAGATGACCACCCA 2256
           |||||||| |          | || ||| |  | | | |||| |||||||   |||  ||
Db     316 ACATCAGCAGCAGCAGCAGCAACAGCAGCAGTTCCAGGCTCAGCAGATGGTCCATGCA 375

Qy    2257 TTTATGAGTCTTGTCTTTGGAGTAGCAGAATCAGAGACACCAGCACCAAGCTCAG 2311
           |    |  ||   |  | | |||||| | ||   | ||||| ||    |||
Db     376 TTTCTGCCTCTGCTGTTTGGAGTGGCAGAATCAGAGACACCAGCACCAAGCTCAG 430

Db     326 GCAGCAGTTCCAAGCAGTAGTGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAG 380


RESULT 14
US-09-902-525-38
; Sequence 38, Application US/09902525
; Parent No US20020164751A1
;  GENERAL INFORMATION:
;       APPLICANT: Recombinant Biocatalysis, Inc.
;       TITLE OF INVENTION: THERMOSTABLE PHOSPHATASES
;       NUMBER OF SEQUENCES: 54
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 4225 Executive Square, Suite 14..
;            CITY: La Jolla
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 92037
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: Windows 95
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/902,525
;            FILING DATE: 09-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/202,681A
;            FILING DATE: 23-Dec 1999
;            APPLICATION NUMBER: 08/666,857
;            FILING DATE: 19 JUN 1996
;            APPLICATION NUMBER: 60/033,752
;            FILING DATE: 19 JUN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Haile, Ph.D., Lisa A.
;            REGISTRATION NUMBER: 38,347
;            REFERENCE/DOCKET NUMBER: 09819/015W01
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 619/678-5070
;            TELEFAX: 619/678-5099
;            TELEX: <Unknown>
;  INFORMATION FOR SEQ ID NO: 38:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 5520 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: Genomic DNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 38:
Db  54 ...GCC GCC 38

  Query Match              1.0%;  Score 44;  DB 4;  Length 5520;
  Best Local Similarity   98.5%;  Pred. No. 1.1;
  Matches  107;  Conservative    0;  Mismatches   1%;  Indels    0;  Gaps    0;

Qy    3974 AGTTTCCACTTCACTACTGAAGACACTTCTTTCATACACTATCTCAACC 4023
           ||  |||   |||||  |  ||||| | |      |||| || | ||  | ||| | | |
Db    3964 AGGGTCCCGATGGACGGGAGAACATAATACGCAACTACCCGCACGACACTTCGTAACC 4023

Qy    4024 GTCTTGACCACAGGGGTAACGCCTTACGACAGGTCCAGAACGCAGACTGTCTC 4073
           | | |   | | |  | |  || | |  | |||  |  ||| ||  | ||| ||
Db    4024 CTCCTGACCAAGGGGCTAACGCCTTACGACAGGTCCAGAACGCCGACTGACTCTC 4073

Qy    3794 CACACAGCCTGGGGGAGGGCCATCAAAGCCATGGCCATCATTTCAAGTCAACTCAGT 3853
           | |     |||||  || ||||||  |   |||      ,    |||||
Db    4084 AGCGACGAGGCCTGGAGGGGCATTAAGGCCGTCAGGCAGGCAAACCAGTAACTCTCAGG 4143

Qy    3854 AAGGACTCCGGCTATGTCATCGCCCTGCGCAG 3885
             ||||   ||  | | |  |  ||| ,|| |
Db    4144 GCGGACATGGGTAAAGACTCCTTCCTCCGCTG 4175
```

ESULT 15
US-09-880-107-2102
Sequence 2102, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921 5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2102
LENGTH: 2038
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION. Genbank Accession No. US20020142981A1 J04029
US 09 880-107-2102

Query Match 1.0%; Score 43.2; DB 10; Length 2038;
Best Local Similarity 46.9%; Pred. No. 1.1;
Matches 135; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy     535 GCTGCAGGAACAGGGCTGGCAGCAGCTAGAACAGCTGTGGGCTCAGCGGCGCTCACAGGC 594
           || || |     ||    ||| || ||||      || |  |   | ||||||     |
Db    1374 GCGGCGGCGGAAGTTTCGGCGGCGGCTACGGAGGCGGAAGCTCCGGCGGCGGAAGCTCCG 1433

Qy     595 CGTGCTCAGTCTTCCACCGTGGCCTCGGAGCCTGTATCAGCCGGCAGCGCCTCGGTCTCCT 654
           | ||      | |   ||    ||||      |          ||||||  ||  || |  ||
Db    1434 GCGGCGGCTACGGGGGCGGAAGCTCCAGCGGCGGCCACGGCGGCAGTTCCAGCGGCGGCT 1493

Qy     655 GCCTCGGATCCAGGCTCCTGTGCCTGGGCTCCAGCCAGGAAGCGATATCTCCAGCGGAG 714
           |   ||  ||||  |||   |||  ||| | |    ||  |||  | |||| |||| |
Db    1494 ACGGTGGTGGCAGTTCCGCGGCGGCGGCTACGGGGGCGGCAGCTCCGGCGGCG 1553

Qy     715 GTCAGCTCTGGGACAGCTGAACACCATTCTCCTAGTGGCCCGGCCCCTGCTCCGGAGACG 774
           |   || | | || ||        | |  |  | ||||  | | | | ||| | ||
Db    1554 GCAGCAGCTCCGGCGGCGGATACGGCGGCGGCAGCAGCTCCGGCGGCGGATACGGCGGCG 1613

Qy     775 ACAGAAGCTACCCTGTGCCCCTGCCCCGCACAGCGGGAGCCCTGGGG 822
           |||   |  ||| |  | |  | ||  | |  | |||  |   || |
Db    1614 GCAGCTCCAGCGGAGGCCACAAGTCCTCCTCTTCCGGGTCCGTGGGCG 1661
```

Search completed: January 10, 2003, 02:19:20
Job time : 291.45 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 07:23:38 ; Search time 30.0438 Seconds
(without alignments)
1938.191 Million cell updates/sec

Title: US-09-803-126-4
Perfect score: 2287
Sequence: 1 MYQSEPGPVFVPVQLEEPPP ... WGPAGEVEPMGEGQAEARDA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2287	100.0	437	22	AAU07400	Human oestrogen-re
2	2217.5	97.0	786	22	AAU07406	Human oestrogen-re
3	1727	75.5	337	22	AAU87214	Novel central nerv
4	1725	75.4	333	22	AAU87523	Novel central nerv
5	1725	75.4	333	22	AAU17193	Novel signal trans
6	1650	72.1	470	21	AAY57895	Human transmembran
7	1543.5	67.5	1337	22	AAU07399	Mouse oestrogen-re
8	306	13.4	56	22	ABB41520	Peptide #9026 enco
9	306	13.4	56	22	ABB25387	Protein #7386 enco
10	306	13.4	56	22	AAM62390	Human brain expres
11	306	13.4	56	22	AAM75200	Human bone marrow
12	306	13.4	56	22	AAM35312	Peptide #9547 enco
13	306	13.4	56	23	ABG44860	Human peptide enco
14	304.5	13.3	2424	22	ABB58924	Drosophila melanog
15	231.5	10.1	1651	22	ABG14648	Novel human diagno
16	227	9.9	2167	22	ABB60869	Drosophila melanog
17	143	6.3	559	22	AAM52317	Human SCaF1 protei
18	143	6.3	559	22	AAG67345	Amino acid sequenc
19	143	6.3	559	22	AAR20494	Human Pablo (pro.a
20	143	6.3	559	22	AAR73399	Human WAVE1 (WAVE/
21	140.5	6.1	216	22	AAB80892	Human ATFx. Homo
22	140.5	6.1	282	22	AAB95859	Human protein sequ
23	140.5	6.1	282	22	AAB80896	Human ATFx with N
24	140.5	6.1	418	22	AAB20496	Human Pablo deltaI
25	136.5	6.0	2129	20	ABB62828	Drosophila melanog
26	134	5.9	425	22	AR701919	Novel human diagno
27	134	5.9	525	20	AAY21625	Ligand binding Pro
28	134	5.9	900	21	AAB42321	Human ORFX ORF2085
29	134	5.9	1217	22	ABG09876	Novel human diagno
30	134	5.9	1239	22	ABG09877	Novel human diagno
31	133	5.8	533	16	AAR72483	Human H-2RIIBP. H
32	132.5	5.8	406	22	ABG27250	Novel human diagno
33	132	5.8	523	14	AAR44468	hPXR-beta1. Homo
34	132	5.8	533	23	AAU78297	Human Retinoid X R
35	131.5	5.7	216	22	ABB50151	Human transcriptio
36	131.5	5.7	1134	23	AAF24341	Human lung specifi
37	129	5.6	694	23	ABB92421	Herbicidally activ
38	128.5	5.6	182	22	AAB80893	Human ATFx leucine
39	128.5	5.6	541	19	AAW37148	Mammalian Ena (Men
40	128.5	5.6	783	19	AAW37151	Mouse neural Mena+
41	128.5	5.6	787	19	AAW37152	Mouse neural Mena+
42	128.5	5.6	802	19	AAW37153	Mouse neural Mena+
43	128.5	5.6	862	22	AAU09139	Mammalian enabled
44	127.5	5.6	1466	20	AAY41091	Rat Tec ;hilin pr
45	127.5	5.6	1466	21	AAB15729	Rat latrophilin.

ALIGNMENTS

RESULT 1
AAU07400
ID AAU07400 standard; Protein; 437 AA.
XX
AC AAU07400;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human oestrogen-regulated unconventional myosin-related protein MRP #1.
XX
KW Human; oestrogen-regulated unconventional myosin-related protein; MRP;
KW auditory; antiatherosclerotic; osteopathic; neuroprotective; oestropic;
KW antiparkinsonian, cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
KW Alzheimer's disease, Parkinson's disease, chromosome 17q25; immunogenic.
XX
OS Homo sapiens.
XX
PN WO200168866-A2
XX
PD 20-SEP-2001.
XX
PF 12 MAR-2001; 2001WO-US08060.
XX
PR 10 MAR-2000; 2000US-188488P.
PR 09 MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Brooks AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS13821.

XX
PT Novel oestrogen-regulated unconventional myosin-related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 1; Fig 4; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen-related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the amino
CC acid sequence of human MRP variant #1 as described in the invention.
XX
SQ Sequence 437 AA;

Query Match 100.0%; Score 2287; DB 22; Length 437;
Best Local Similarity 100.0%; Pred. No. 2.7e-197;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVQVPFSPVPMPVQPQPPPYAPLPPIFPFPALAPLPIMPAHSPPFMLPFPFPFSIPAAV 60
Db 1 MVQVPFSPVPMPVQPQPPPYAPLPPIFPFPALAPLPIMPAHSPPFMLPFPFPFPAAV 60

Qy 61 APPPKAPLQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 120
Db 61 APPPKAPLQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 120

Qy 121 PPQPVPPNQVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 180
Db 121 PPQPVPPNQVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 180

Qy 181 PPPSPSPPPSQNFPPPMMPPPPPPPLFVPLPPLPTTPPPSVPFPPIVVAAPPNWANYFSPPPP 240
Db 181 PPPSPSPPPSQNFPPPMMPPPPPPPLFVPLPPLPTTPPPSVPFPPIVVAAPPNWANYFSPPPP 240

Qy 241 VSPPPPPPPVVLLAVSHPGLPLPLPVTQGPGLPPPQLKPLPCSVSPAPVLGVPCPGGGTLELS 300
Db 241 VSPPPPPPPVVLLAVSHPGLPLPLPVTQGPGLPPPQLKPLPCSVSPAPVLGVPCPGGGTLELS 300

Qy 301 LPSPQLVLHTAPAPAIPALVELPLNELPPQCPVYIALPCYITPNCPLLSPHPGPLIPLLP 360
Db 301 LPSPQLVLHTAPAPAIPALVELPLNELPP- ------ALPCYITPNCPLLSPHPGPLIPLLP 354

Qy 361 VNTPPPWQLPQAQPGGLPPAPPIVPAAAPPPPPPPEQPPQPGWHPQQLSNSPPQLAPWPP 420
Db 355 VPTPPPWQLPQAQPGGLPPAPPIVPAAAPPPPPPPEQPPQPGWHPQQLSNSPPQLAPWPP 414

Qy 421 ASPVPPMSPGQAPAPPA 437
Db 415 ASPVPPMSPGQAPAPPA 431

RESULT 2
AAU07406
ID AAU07406 standard; Protein; 786 AA.
XX
AC AAU07406;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human oestrogen regulated unconventional myosin related protein MRP #2.
XX
KW Human; oestrogen regulated; unconventional myosin related protein; MRP;
KW auditory; antiatherosclerotic; cytopathic; neuroprotective; nootropic;
KW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
KW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic.
XX
OS Homo sapiens.
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHE) SCHERING AG.
XX
PI Brooks AP, Deng CC, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS13823.
XX
PT Novel oestrogen regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 1, Fig 4, 108pp, English.
XX
CC The invention relates to an isolated oestrogen regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the amino
CC acid sequence of human MRP variant #2 as described in the invention.
XX
SQ Sequence 786 AA;

Query Match 97.0%; Score 2217.5; DB 22; Length 786;
Best Local Similarity 97.5%; Pred. No. 1.1e-190;

```
 Matches  426;  Conservative    0;  Mismatches    0;  Indels   11;  Gaps    1;

Qy       1 MYQSRPGPVPVPVQPSPPPKAPLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MYQSRPGPVPVPVQPSPPPKAPLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60

Qy      61 APRPKAPLQLGPSSSIKPKQGPLLDLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 APRPKAPLQLGPSSSIKPKQGPLLDLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRC 120

Qy     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLPLPKPVPYPRENFSHPYYLPLLCEQILK 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLPLPKPVPYPRENFSHPYYLPLLCEQILK 180

Qy     181 DTFSESCIRIGQNEPPKMKDLLGGLEVDLDSLTTTEDSVKKPIVVAARDNWANYFSPFPP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DTFSESCIRIGQNEPPKMKDLLGGLEVDLDSLTTTEDSVKKPIVVAARDNWANYFSPFPP 240

Qy     241 VSGFSGSDVQLLAVSHPGLPLLKVTQGPGLRPDQLKILCSYSFAEVLGVECPGGSTLELS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VSGFSGSDVQLLAVSHPGLPLLKVTQGPGLRPDQLKILCSYSFAEVLGVECPGGSTLELS 300

Qy     301 LKSEQLVLHTAPAPAIPALVELPLNELKKDSGYVIALRSYITDNCSLLSPHPGDLIFLLP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 LKSEQLVLHTAPAPAIPALVELPLNELKKDSGYVIALRSYITDNCSLLSPHPGDLIKLLP 360

Qy     361 VATLEPGWQPGSAGGRRGGLPPADIVQPAAAPDFSPSKEQRSGWHKGQLSNGEPGLAPWDP 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 VATLEPGWQPGSAGGRRGGLPPADIVQPAAAPDFSPSKEQRSGWHKGQLSNGEPGLAPWDR 420

Qy     421 ASEVRKMGPGQAEARPA 437
           |||             |||
Db     421 ASE-----------RPA 426


RESULT 3
AAU87214
ID   AAU87214 standard; Protein; 337 AA.
XX
AC   AAU87214;
XX
DT   05-JUN 2002  (first entry)
XX
DE   Novel central nervous system protein #124.
XX
KW   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW   acquired immunodeficiency; virus; dysphagia; gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200155318 A2.
XX
PD   02-AUG-2001
XX
PF   17 JAN-2001; 2001WO-US01332.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180004.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   14-AUG-2000; 2000US-0225779.
PR   18-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234273.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02 OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US 0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
```

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 21-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250422.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-591617/65.
DR N-PSDB, AAK13544.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical disorders and used as
PT food additives or preservatives.
XX
PS Claim 9; SEQ ID No 732; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in hair care. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 76.5%; Score 1727; DB 22; Length 837;
Best Local Similarity 99.4%; Pred. No. 5.2e-147;
Matches 334; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 102 APPLPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSSVPPYTTWWKLPLRRRVPYP 161
Db 2 ASALPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSSVPPYTTWWKLPLRRRVPYP 61

Qy 162 RENICHFYYLRLLCEQLRDTHSRVTKQNRNKFMKELSFYEMPEFMPSEAAEQVKSLRDLA 221
Db 62 RENFSHPYYLRLLCEQLRDTHSRVTKQNRNKFMKELSFYEMPEFMPSEAAEQVKSLRDLA 121

Qy 222 RIVVAARDNWANYPSRPFPPVSCPWDYVQLLAVSHRLRLLKVTLRSLRPLQRLKLDSY 281
Db 122 RIVVAARDNWANYPSRPFPPVSGEDSDVQLLAVSHRLRLLKVTLRSLRPLQRLKLDSY 181

Qy 282 SPAEVLGVECRGGSTLRLSLKSEELIVIHTAPAPAIPALVRLFLNRLFPLGTYVIAERSY1 341
Db 182 SPAEVLGVECRGGSTLRLSLKSEELIVIHTAPAPAIPALVRLFLNRLFPLGTYVIAERSY1 241

Qy 342 TDNCSLLSFHRGDLIKLLPVATLEPGWQPGSAGGRSGLFPADIVQTAAATTFQFSREQRS 4 1
Db 242 TDNCSLLSFHRGDLIKLLPVATLEPGWQPGSAGGRSGLFPADIVQTAAATTFQFSREQRS 4 1

Qy 402 GWHPQQLTGFFQLARWTPATPVPPMSPQAPAPPA 437
Db 302 GWHPQQLTGFFQLARWTPATPVPPMSPQAPAPPA 337

RESULT 4
AAU87523
ID AAU87523 standard; Protein; 333 AA.
XX
AC AAU87523;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #433.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX

PR 02-AUG-2000;
XX
PR 17-JAN-2001; 2001WO-US01332
XX
PR 31-JAN-2000; 2000US-0180628.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186250.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190122.
PR 18-APR-2000; 2000US-0198515.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215136.
PR 07-JUL-2000; 2000US-0216047.
PR 07-JUL-2000; 2000US-0216980.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217488.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220064.
PR 26-JUL-2000; 2000US-0220064.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226079.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226869.
PR 22-AUG-2000; 2000US-0227180.
PR 23-AUG-2000; 2000US-0227884.
PR 30-AUG-2000; 2000US-0229324.
PR 01-SEP-2000; 2000US-0229387.
PR 01-SEP-2000; 2000US-0229943.
PR 01-SEP-2000; 2000US-0229944.
PR 01-SEP-2000; 2000US-0229945.
PR 05-SEP-2000; 2000US-0230609.
PR 05-SEP-2000; 2000US-0230613.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231343.
PR 08-SEP-2000; 2000US-0231344.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0241968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-581633/65.

```
DR   N-PSTR; ABK43873.
XX
PT   New isolated nucleic acid encoding a protein for diagnosing,
PT   prevention, treating or ameliorating medical conditions and used as
PT   food additives or preservatives
XX
PS   Claim 7; SEQ ID NO 1041; 827pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule (I) encoding a
CC   novel central nervous system protein. (I) and polypeptides (III) encoded
CC   by (I), are used to treat a medical conditions and in diagnosis of a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC   amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC   e.g. Acquired immunodeficiency virus (AIDs) and fungi, ocular disorders
CC   e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC   adenocarcinomas and irritable bowel syndrome, reproductive system
CC   disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC   and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC   leukaemia, disorders involving neovascularisation e.g. malignancies,
CC   respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC   acute kidney failure and blood related disorders e.g. myocardial
CC   infarction. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities, fat content, lipid, protein,

  Query Match             75.4%;  Score 1725;  DB 22;  Length 333;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-147;
  Matches  333;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MPLPEPVPGTLSAERRFCLTQFVPPLSWTPLLAFVLLFWFLFLRPEVFPPPEN 154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MPLPEPVPGTLSAERRFCLTQFVPPLSWTPLLAFVLLFWFLFLRPEVFPPPEN 154

Qy    155 PCHPVTLRLLCPELRLTPLRPIRLLNERRFMRLLLSLLEVLLSLTTELSVFFPIV 334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 PCHPVTLRLLCPELRLTPLRPIRLLNERRFMRLLLSLLEVLLSLTTELSVFFPIV 334

Qy    335 VAARDNWANYFCPPFPVCEDSSLVLLAVSMFGLFLLFVTVAFCLKLQLRPLLTCYSPA 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 VAARDNWANYFCPPFPVCEDSSLVLLAVSMFGLFLLFVTVAFCLKLQLRPLLTCYSPA 494

Qy    495 EVLTVFCPLGLPLSCLVDTATFAFALELVFLNELFFTSGYYVALRSYIIDN 644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 EVLTVFCPLGLPLSCLVDTATFAFALELVFLNELFFTSGYYVALRSYIIDN 644

Qy    645 FLLSTANLPLLPVATLEPGWQPSAQGPSSLFPAFIVAAAAPPSPPPEAWSWH 804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    645 FLLSTANLPLLPVATLEPGWQPSAQGPSSLFPAFIVAAAAPPSPPPEAWSWH 804

Qy    805 KSLANSPPGLARWDRASEVFFMGEGQAEAPPA 837
          |||||||||||||||||||||||||||||||
Db    805 KSLANSPPGLARWDRASEVFFMGEGQAEAPPA 837


RESULT 5
AAU17193
ID   AAU17193 standard; Protein; 333 AA.
XX
AC   AAU17193;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Novel neural transduction pathway protein, Seq ID 758.
XX
KW   Neurological disorder, psychiatric, dermatological, immunosuppressive, tumour,
KW   anti-inflammatory, anti-HIV, antibacterial, antiinflammatory; cancer;
KW   immune system disorder, rheumatoid arthritis, inflammatory bowel d;
KW   organ transplant rejection; infection; hepatitis C; blood disorder;
KW   sickle cell anaemia, hyperproliferative disorder; Gaucher's disease;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW   cardiovascular; respiratory; wound healing; eye disease; Addison's disease;
KW   reproductive system; gastrointestinal; liver disorder; AIDS;
KW   acquired immune deficiency syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200154733-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01312.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184684.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   19-APR-2000; 2000US-0198124.
PR   19-MAY-2000; 2000US-0205515.
PR   [illegible]-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   06-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
```

PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 21-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 26-SEP-2000; 2000US-0235934.
PP 27-SEP-2000; 2000US-0235936.
PP 27-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PP 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237045.
PP 13-OCT-2000; 2000US-0239947.
PR 13-OCT-2000; 2000US-0240569.
PP 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PP 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PP 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 08-NOV-2000; 2000US-0249207.
PP 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PP 17-NOV-2000; 2000US-0249211.
PP 17-NOV-2000; 2000US-0249212.
PP 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PP 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PP 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249264.
PP 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PP 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0250150.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 05-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR N-PSDB; AAB27110.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID NO 758, 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 75.4%; Score 1725; DB 22; Length 333;
Best Local Similarity 100.0%; Pred No. 7.7e-147;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 LPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSTTSTPWPLPTPPPVFYPREN 164
Db 1 LPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSTTSTPWPLPTPPPVFYPREN 60

Qy 165 PSHPYYLPLLCPQTLRDTFSESCTKISQNERPPMKPDLLGGLEVDLDSLTTTEDSVPPRTV 224
Db 61 PSHPYYLPLLCPQTLRDTFSESCTKISQNEPPPMKPDLLGGLEVDLDSLTTTEDSVPPRTV 120

Qy 225 VAARDNWANYFSRFFPVSGESGSDVQLLAVSHRGLPTLEVTQGLSLPPPQLKTLCSYSPA 284
Db 121 VAARDNWANYFSRFFPVSGESGSDVQLLAVSHRGLPTLEVTQGLSLPPPQLKTLCSYSPA 180

Qy 285 EVLGVECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKFDSGYVIALPSYITDN 344
Db 181 EVLGVECRGGSTLELSLKSFQLVLHTAPARAIEALVELFLNELKFDSGYVIALPSYITDN 240

Qy 345 CSLLSFHRGDLIKLLPVATLEPGWQFGSAGGPSGLPPADIVQPAAAPDFSPSKLQRSGWH 404
Db 241 CSLLSFHRGDLIKLLPVATLEPGWQFGSAGGPSGLPPADIVQPAAAPDFSPSKLQRSGWH 300

Qy 405 KGQLSNGEPGLAPWDRASEVPKMGEGQAEARPA 437
Db 301 KGQLSNGEPGLAPWDRASEVPKMGEGQAEARPA 333

RESULT 6
AAY57895
ID AAY57895 standard; Protein; 470 AA.
XX
AC AAY57895;
XX
DT 28-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN-19.
XX
KW Human; transmembrane protein, HTMPN, diagnosis, immunospecific;
KW antiproliferative; neuroprotective, immune disorder,
KW reproductive disorder, smooth muscle disorder, neurological disorder;
KW gastrointestinal disorder; developmental disorder,
KW cell proliferative disorder.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11904.
XX
PR 29-MAY-1998; 98US-0087060.
PR 31-JUL-1998; 98US-0091674.
PR 27-OCT-1998; 98US-0105954.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Edelman J, Patterson C, Corgone GA, Kaser MR, Baughn MR,
PI Au Young J;
XX
DR WPI; 2000-072695/06.
DR N-PSDB; AAZ56716.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders
XX
PS Claim 1; Page 125-126; 229pp; English
XX
CC AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN 1 to HTMPN 79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or
CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.
XX
SQ Sequence 470 AA;

  Query Match   72.1%; Score 1650; DB 21; Length 470;
  Best Local Similarity   84.0%; Pred. No. 7.1e-140;
  Matches 385; Conservative   8; Mismatches   34; Indels   22; Gaps   5;

Qy   47 MLTPSPLKTTFAVARPFAPLQLSPSSSIFPFQSPLLLFSQFLFIAHTFPFPFAPFIP 106
           ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MLLPSPLRTTFAVARSFAPLQLSPSSSIFPFQSPLLLFSQFLFIAHTFPFPFAPFIP 60

Qy   107 LSVTQWLLAFLSLFSVCFSYTSIFWFLFLFFEVFYPFSNFS 166
           ||:||:|||||||||||||||||||||||||||||||||||||
Db   61 LPEDPSTLSAERRCYTQPVFDQAVSIQLLAFSSSVCFSYTSIFWFLFLFFEVFYPFSNFS 120

Qy   167 HPYYLRLLCEQILRDTFSESCIRISQNERRRMKDLLRGLEVGLLSGLTTTEDSVKERIVVA 226
           ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   121 HPYYLRLLCEQILRDTFSESCIRISQNERRRMKDLLRGLEVGLLSGLTTTEDSVKERIVVA 180

Qy   227 AENWANYFSRFFFVSGESSGDVQLLAVSHRGLRLLEVTQHFRDITQLRKLCTTNFAEV 286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AENWANYFSRFFFVSGESSGDVQLLAVSHRGLRLLEVTQHFRDITQLRKLCTTNFAEV 240

Qy   287 LGVECPGGSTLELSLKSEQLVLHTAPAPAIPALVELFLNELFKLGNYVIALRSYITNCT 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 LGVECPGGSTLELSLKSEQLVLHTAPAPAIPALVELFLNELFKLGNYVIALRSYITNCT 300

Qy   347 LLSFHRGDLIKLLPVAILEPG-----W-----QPSAAQPPTLLPPAILVLVAAA 396
           ||||||||||||||||||||| | | | | | | | |
Db   301 LLIFHRGDLIKLLFVC--HPGAFLATWLTRCIFKTLSCRHAAAFKSKLPLLQGAEEWLA 358

Qy   397 LLPSPPPLPSSWHPSLNSFL-----SPAPWSPASPVEP 436
           | | ||| || || | | | | |||
Db   359 LGSAVQPGTRAG-SVGGGLKGEELGRGISRGRAGLRLRK 396

RESULT 7
AAU07399
ID AAU07399 standard; Protein; 1337 AA.
XX
AC AAU07399;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mouse oestrogen regulated unconventional myosin related protein, MMRP.
XX
KW Mouse; oestrogen-regulated unconventional myosin related protein, MMRP;
KW fertility, antiatherosclerotic, osteopathic, breast cancer; neutropic;
KW antiprotozoan, cytostatic, oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
KW Alzheimer's disease; Parkinson's disease; chromosome 11; immunosuppressive.
XX
OS Mus sp.
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHE) SCHERING AG.
XX
PI Brinks AF, Berg OG, Fehenyi GM;
XX
DR WPI; 2001-600747/68.
DR N-PSDB; AAS13813; AAS13820.
XX
PT Novel oestrogen regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 1; Fig 1; 108pp; English.
XX
CC The invention relates to an isolated oestrogen regulated unconventional
CC myosin related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and

CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen-related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the amino
CC acid sequence of mouse MRP as described in the invention.
XX
SQ Sequence 1337 AA;

Query Match 67.5%; Score 1543.5; DB 22; Length 1337;
Best Local Similarity 71.4%; Pred. No. 1.2e-129;
Matches 308; Conservative 34; Mismatches 77; Indels 15; Gaps 3;

Qy 1 MYQSRPGFVEVDVQESEPFKAFLPKIDPKDFEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60
Db 914 MYQSFPGPVAVDVLPTFPIPTFQKYNDPKDFEALAKLGINGVH-LPLSTSPNQGKSSPPAV 972

Qy 61 APPPKAPLQLGPSGSIKFFQQGPLLDLFG---QKLPIAHTPPPPPAPPLPLPEDPGTLSAE 117
Db 973 VPRPKARPRLEFSLSIQEKQGPLPDLFGPCSPNPPTAPAPPPPPALPPPLSGEPKTPSVE 1032

Qy 118 RRGLTQFVERQGVSTQLIAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSHPYYLRLLCEQ 177
Db 1033 SHALTEPMEDKNISTKLLVPSGSVCFSYANAPWKLFLPFEVFYPRFNFSHPYCLSLLCQQ 1092

Qy 178 ILRDTFSESCIRISQNEPPFMKDLIGGLEVDLDSLTTTEDSVKKRIVVAARDNWANYFSR 237
Db 1093 ILPDTFTESCIRISQDEPHFMKGLLGDLEVSLETLDIVEDSIKKRIVVAARDNWANYFSR 1152

Qy 238 FFPVSGESGSDVQLLAVSHRGLPLLKVTQGPGLRPDQLKILCSYSFAEVLGVECRGGSTL 297
Db 1153 IFPVSGESGSDVQLLGVSHRGLPLLKVTQSPSFHLDQLKTLCSYSYAEVLTVQCRGRSTL 1212

Qy 298 ELSLKSEQLVLHTAPARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIK 357
Db 1213 ELSLKNEQLILHTAWARAIKAMVDLFLSELRKDSGYVIALRSYITDDNSLLSFHRGDLIR 1272

Qy 358 LLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAR 417
Db 1273 LLPVTALEPGWQFGSAGGRSGLFPDDVVQPAAAPDLSFSLGKRNSWQR---------KSK 1323

Qy 418 WDPASEVFPNGE 429
Db 1324 LGPAQEVRKTEE 1335

RESULT 8
ABB41520
ID ABB41520 standard; Peptide; 56 AA.
XX
AC ABB41520;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #9026 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 34155; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 56 AA;

Query Match 13.4%; Score 306; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 4.3e-20;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 368 WQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAPWDPASE 423
Db 1 WQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAPWDPASE 56

RESULT 9
ABB25387
ID ABB25387 standard; Protein; 56 AA.
XX
AC ABB25387;
XX
DT 22-JAN-2002 (first entry)
XX
DE Protein #7386 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension, cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

```
XX
DR   WPI; 2001 488899/53.
XX
PT   Single exon nucleic acid probes for analysing gene expression in human
PT   hearts
XX
PS   Claim 15; SEQ ID No 27157; 538pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   AR441 [* AR441205]). The present sequence is a protein encoded by one such
CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and congenital heart disease
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   56 AA;

  Query Match             13.4%;  Score 306;  DB 22;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-20;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     368 WQFSAAFRFSQLFFALIVQFAAAIFSFSKPQRSGWHKGQLSNGFPGLARWDRASE 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 WQFSAAFRFSQLFFALIVQFAAAIFSFSKPQRSGWHKGQLSNGFPGLARWDRASE 56


RESULT 10
AAM62390
ID   AAM62390 standard; Protein; 56 AA.
XX
AC   AAM62390;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 34495.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   28-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488446/52.
XX
PT   Single exon nucleic acid probes for analysing gene expression in human
PT   brains
XX
PS   Example 4; SEQ ID NO: 34495; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. The, can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of neurological
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention.
XX
SQ   Sequence   56 AA;

  Query Match             13.4%;  Score 306;  DB 22;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-20;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     368 WQFSAAFRFSQLFFALIVQFAAAIFSFSKPQRSGWHKGQLSNGFPGLARWDRASE 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 WQFSAAFRFSQLFFALIVQFAAAIFSFSKPQRSGWHKGQLSNGFPGLARWDRASE 56


RESULT 11
AAM75200
ID   AAM75200 standard; Protein; 56 AA.
XX
AC   AAM75200;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 35506.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   28-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow
XX
PS   Example 4; SEQ ID NO: 35506; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   56 AA;

  Query Match             13.4%;  Score 306;  DB 22;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-20;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     368 WQFSAAFRFSQLFFALIVQFAAAIFSFSKPQRSGWHKGQLSNGFPGLARWDRASE 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 WQFSAAFRFSQLFFALIVQFAAAIFSFSKPQRSGWHKGQLSNGFPGLARWDRASE 56
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 WQFGSAGGRSGLFPADIVQFAAAFDFSFSKFQPSGWHKGQLSNGEPGLAPWDRASE 56

RESULT 12
AAM35312
ID   AAM35312 standard; Protein; 56 AA.
XX
AC   AAM35312;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #9349 encoded by probe for measuring placental gene expression.
XX
KW   Probe, microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR.
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 35581; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   56 AA;

  Query Match             13.4%;  Score 306;  DB 22;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-20;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        368 WQFGSAGGRSGLFPADIVQPAAAFDFSFSKEQPSGWHFGQLSNGEPGLAPWDRASE 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 WQFGSAGGRSGLFPADIVQPAAAFDFSFSKEQPSGWHKGQLSNGEPGLARWDRASE 56


RESULT 13
ABG44860
ID   ABG44860 standard; Peptide; 56 AA.
XX
AC   ABG44860;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human peptide encoded by genome-derived single exon probe SEQ ID 34525.
XX
KW   Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis, Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US00665.
XX
PR   04-FEB-2000; 2000US-180312P.
PR   26-MAY-2000; 2000US-207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-234687P.
PR   27-SEP-2000; 2000US-236359P.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, ...
PT   measure gene expression in human lung samples -
XX
PS   Claim 27; SEQ ID No 34525; 434pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequence
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human lung derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a peptide/protein
CC   encoded by a single exon probe of the invention.
CC   Note: The sequence data for this patent did not form part
```

```
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   56 AA;

  Query Match             13.4%;  Score 306;  DB 23;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-17;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      368 WQPGSAQGRSGLPPAGIVQPAAAAPDPSPSRE 403
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 WQPGSAQGRSGLPPAGIVQPAAAAPDPSPSRE 56


RESULT 14
ABB58924
ID   ABB58924 standard; Protein; 2424 AA.
XX
AC   ABB58924;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 3564.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   24-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03027.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 3564; 31pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL03176-ABL39511), expressed DNA
CC   sequences (ABL21849-ABL16179) and the encoded proteins
CC   (ABB57747-ABB72672).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2424 AA;

  Query Match             13.3%;  Score 304.5;  DB 22;  Length 2424;
  Best Local Similarity   33.7%;  Pred. No. 1.1e-17;
  Matches   93;  Conservative   38;  Mismatches   96;  Indels   41;  Gaps   10;

Qy      99 QRLPLAG-TPPPPAPPLPLPEPPGTLSAERRCLT-PVEL-GVSTQLLAPSGS----VCF 143
           ||| | | |||   | |||     |              || : :|  |||     |
Db    2173 RRLDGMRMTTTTTVIVPAPVPPPGGTSPG        - LGVVSGVLMGGGGGGGTAGL 2218

Qy     144 SYTGTPWKLFLPKFVPYPPPNPSHPYYLPLLSTQLLPNTPGPS-TTKTQLNEPPPMPLL 202
           :|   |||| :||||| | |     |  | :|  ||: | |  : |:|:  |:    ::
Db    2219 TYNPVPWPLPVPPPVPQPHPPPGPPVAIIIPAQVITVVPGVT-TTKTTNLPPSNANMT  277

Qy     203 GGLEVDLDSLTTTEDS-------  ---  VKKPIVVAAPLNWANYPSKPPPVSKKS 247
           |  | :|:|:    |             ||: :|  |:: : :::  | | :
Db    278 HPHP-GIPLNPPSSA-AIIGIGLQ-AIVPPHQTMAPL AVIPAPIPAL PAIPI 327

Qy     248 DVQLLAVSHRGLPLLKVTQGPGLKPLQLKPLQSYAPAPV LSVEVPPPSTLEILGLKS PL 307
           || :: ||| || | :           | | :: : || |:   :    : :|:|:: :
Db    328 DVSIMGVGHSGLYLAR       RLADTLIVVQAISPGEIPAAVTLIPIAANQLNLANIKH 380

Qy     306 LVLHTARARAIEALVELFLNELKK 329
           | || ||| ||:::|  |: | :|
Db    2392 LALHAARAAAIQSMVTTFVQEFRK 2415


RESULT 15
ABG14648
ID   ABG14648 standard; Protein; 1651 AA.
XX
AC   ABG14648;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #14639.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   28-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-683FT/74.
DR   N-PSDB; AAG78835.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 45007; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (II) is useful in gene therapy techniques
CC   to restore normal activity of (II) in treating disorders involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
```

C diagnostic amino acid sequences of the invention.
C Note: The sequence data for this patent did not appear in the printed
C specification, but was obtained in electronic format directly from WIPO
C at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1651 AA;

Query Match 10.1%; Score 231.5; DB 22; Length 1651;
Best Local Similarity 24.3%; Pred. No. 2.4e-11;
Matches 110; Conservative 59; Mismatches 143; Indels 141; Gaps 19;

```
Qy      31 EALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRPKAPLQL------------GPSSS-- 75
           | | :   ||  : ||                 ||  :   |:|:             |:|
Db     512 ERLRRTYANGVRAEPPTWLEL------QAVKSKKHIPIQVILATGESLTVPVDSASTSRE 565

Qy      76    -IKEKQGPLLDLFGQKLPIA------------HTPPPPPAPPLPLPEDPGTLSAER 118
              |  ||| | |  |  | :|               |                 : |
Db     566 MCMHIAHKQG LSDHLGFSLQVAVYDKFWSLGSGRDH----------------MMDAIA 607

Qy     119 RCLTQPVFDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYP-RENFSHPYYLRLLCEQ 177
           ||   |  :::| | :               :||::: ||| | |  ..   |   |:  |
Db     608 RC-EQMAQERGESQR . . . . . QSPWPIYFRKEFFTPWHDSREDPVSTELIYRQ 653

Qy     178 ILRDTFS--------ESCIRI-------- -SQNERPKMKDLLGG-------------- 204
           :||  :|        |  : :         :  | : :::||
Db     654 VLRGVWSGEYSFRKEEELVELLARHCYVQLGASAESKAVQELLPSCIPHKLYRTKPPDRW 713

Qy     205 --------LEVDLDSLTTTEDSVKKRIVVAARDNWANYFSPFFPVSGESGSDVQ----LL 252
                   :          |  :|:::::| |||  |   ||| | |   ||  :     :|
Db     714 ASLVTAACAKAPYTQKQVTPLAVREQVVDAARLQWPLLFSRLFEVITLSGPRLPKTQLIL 773

Qy     253 AVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGV ---ECPGGSTLELSLKSEQLVL 308
           ||:  :||  |              | |:|   || ||:|:   | .:|  | ||   |
Db     774 AVNWKGLCFL---------DQQEKMLLELSFPEVMGLATNREAQGGQRLLLSTMHEEYEF 824

Qy     309 HTARARAIEALVELFLNELKKDSGYVIALRS-YITDNCSLLSFHRGDLIKLLPVATL--E 365
           :  : ||  || |||  ||: | : :||:    ||. :||:| :|||: |      |
Db     825 VSPSSVAIAELVALFLEGLKERSIFAMALQDRKATDDTTLLAFKKGDLLVLTKKQGLLAS 884

Qy     366 PGWQFG--SAGGRSGLFF    ----ADIVQPAA 389
             | |      |::|| |          : :|:|
Db     885 ENWTLGQNDRTGKTGLVPMACLYTIPTVTKPSA 917
```

Search completed. January 10, 2003, 02:17:48
Job time : 33.0438 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

Run on: January 10, 2003, 03:17:59 ; Search time 15.534 Seconds
(without alignments)
2704.441 Million cell updates/sec

Title: US-09-803-126-4
Perfect score: 2287
Sequence: 1 MYQSKPGFVPVPVQPSPPPK ..... WDEAGEVPKMGRGQAEARPA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	689.5	30.1	3511	2	A59295	unconventional myo
2	688	30.1	1783	2	T42386	unconventional myo
3	673.5	29.4	3533	2	A59266	hypothetical prote
4	204.5	8.9	2810	2	T22298	myosin VIIa, long
5	199	8.7	2175	2	A59255	myosin VIIa - mous
6	196	8.6	2215	2	T30870	hypothetical prote
7	182	8.0	2038	2	T26988	structural protein
8	148.5	6.5	555	2	T30349	proline-rich prote
9	138	6.0	449	2	S16748	myosin VII-like pr
10	137.5	6.0	2121	2	A59233	proline-rich prote
11	135.5	5.9	534	2	S21961	probable cytoskele
12	135	5.9	1420	2	T37781	formin - mouse
13	134	5.9	1468	2	S11515	exensin homolog F
14	133	5.8	699	2	T05225	formin isoform 1V
15	132	5.8	1206	2	S24402	unknown protein [i
16	130.5	5.7	1201	2	G86441	retinoid X recepto
17	129	5.6	533	2	S37781	homeotic protein C
18	129	5.6	1426	2	T30817	hypothetical prote
19	128	5.6	727	2	C84584	extensin-like prot
20	128	5.6	1188	2	S49915	extensin-like prot
21	127.5	5.6	228	2	S63504	extensin - common
22	127.5	5.6	463	2	T12698	CL1AA protein - ra
23	127.5	5.6	1466	2	T17138	CL1BA protein - ra
24	127.5	5.6	1471	2	T17149	CL1AB protein - ra
25	127.5	5.6	1510	2	T17145	CL1BB protein - ra
26	127.5	5.6	1515	2	T17156	extensin - Volvox
27	127	5.6	464	2	S22697	hypothetical prote
28	126.5	5.5	876	2	T49801	T20H2.9 protein -
29	126.5	5.5	1137	2	A86335	gag polyprotein -
30	126	5.5	512	1	FOMVGS	proline-rich prote
31	125.5	5.5	379	2	T05441	extensin-like prot
32	125.5	5.5	379	2	D85257	protein-tyrosine-p
33	125	5.5	1494	2	T14355	prpL2 protein - hu
34	124.5	5.4	403	2	S52796	neural-derived rec
35	124.5	5.4	625	2	S71930	hypothetical prote
36	124.5	5.4	731	2	T04455	probable Pro kinas
37	124.5	5.4	760	2	F86387	protein T5E21.13 [
38	124	5.4	1776	2	S86280	glyceraldehyde 3-p
39	123.5	5.4	440	2	I49681	extensin homolog T
40	123.5	5.4	760	2	T06231	actin depolymeriza
41	123	5.4	505	2	S70273	proline-rich prote
42	123	5.4	577	2	T09024	Abl substrate enha
43	123	5.4	684	2	A56154	hypothetical prote
44	123	5.4	708	2	D96711	hypothetical prote
45	123	5.4	1006	2	G86292	

ALIGNMENTS

RESULT 1
A59295
unconventional myosin-15 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59295
R;Liang, Y.; Wang, A.; Belyantseva, I.A., Anderson, D.W.; Probst, F.J.; Barber, T.D.; Mil
an, T.P.; Fridell, R.A.
Genomics 61, 243-258, 1999
A;Title: Characterization of the human and mouse unconventional my . . . vv genes responsi
A;Reference number: A59266; MUID:20021762; PMID:10552926
A;Accession: A59295
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3511 <LIA>
A;Cross-references: GB:AF144095; NID:g6224684; PIDN:AAF05904.1; PID:g6224685
C;Genetics:
A;Gene: MGI:Myo15
A;Cross-references: MGI:1261811
A;Map position: 11:33.9
C;Superfamily: myosin motor domain homology
F;1209-1871/Domain: myosin motor domain homology <MMO>

Query Match 30.1%; Score 689.5; DB 2; Length 3511;
Best Local Similarity 33.6%; Pred. No. 7.4e-39;
Matches 184; Conservative 72; Mismatches 145; Indels 147; Gaps 15;

Qy 4 SPPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGFTPAVAPR 63

Db 2405 SPPQLPERPLAPEARPKTVVG----------TGPPAKPVLVRPTPQSWAISPVAPA 2510

Qy 64 PKAP----------LQLGPS----------SSIKE-KCGF 92

Db 2511 PKIPSKPVAVPILAQDWTAPESISASPELVRYSTLNSEHFPQPTQQIKSTIPQYPQPWA 2570

Qy 93 ----------LLDLFGQKLPIAHTPPPPPAPPLFLFLDPSTLGAPP 118

Db 2571 GHPEARRTDGGKVFPPPPDPHEEALMILKGQKTQLAVVPGTQVSREAVAMVKPVI SAPP 2629

Qy 119 PGL----------TQPVEDQGVSTQLLAPSGSVCFSYTGTFWKLPLPKEVFYFRENFSHP 168

Db 2630 PCMGPTPVQPSRSLEPPEDP-VQTQLHRLVNPNFYGYQNIFWKIPLERFVFYFRDVYGHP 2688

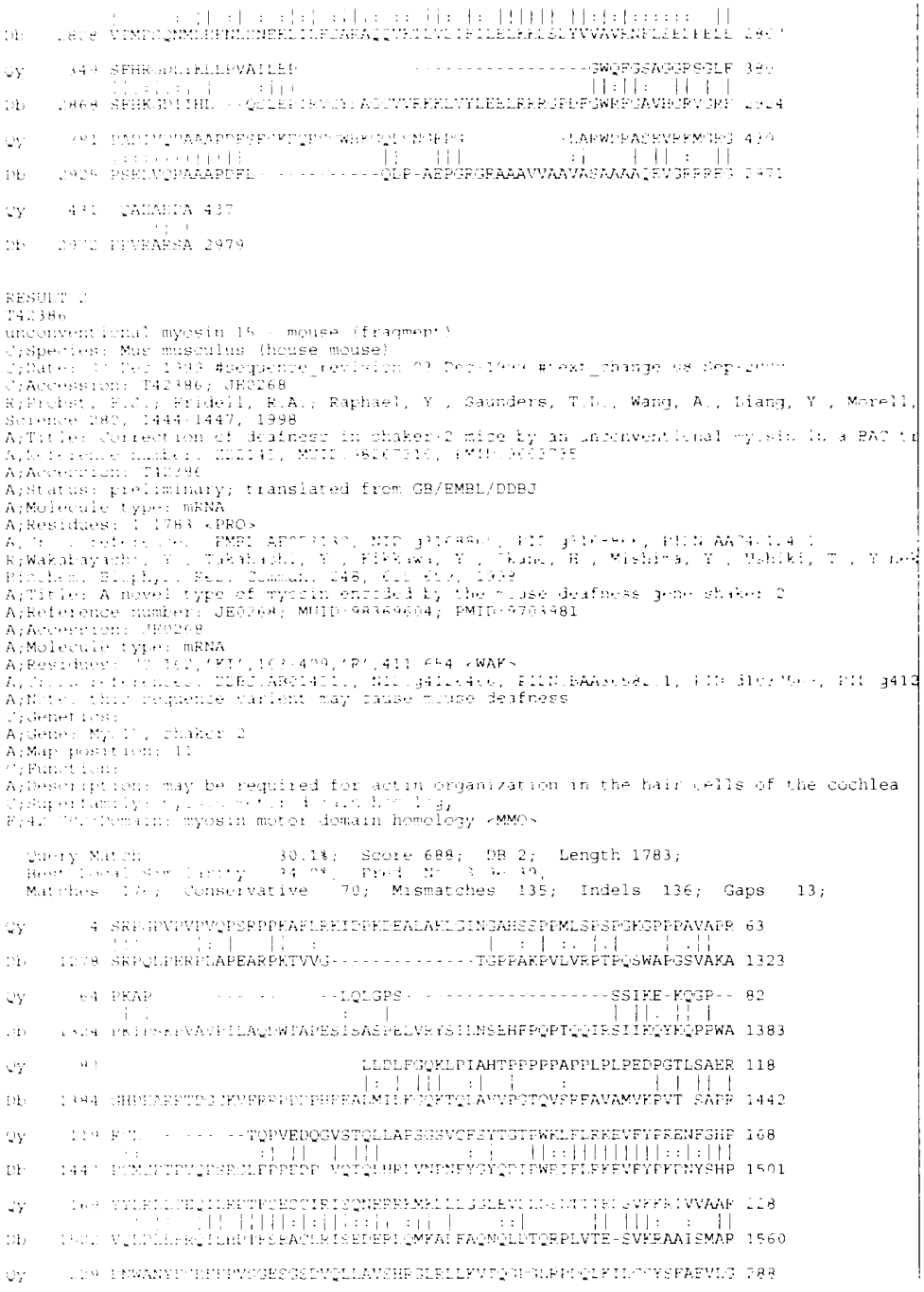
Qy 169 YYLRLLCEQILRDTPSESCIRISQNEFFYMKDLLQGLEVDLDSLITTEDSVKKRIVVAAR 228

Db 2689 VQLDLLFPQIIHDTFSEACLRISEDERLQMKALFAQNQLDTQPPLVTE SVKRAAISMAR 2747

Qy 229 DNWANYFSRFFPVSGESGSDVQLLAVSHPGLPLLKVTQGPGLRPDQLKIICSYGFAEVIG 288

Db 2748 DSWEIYFSRLFPAMGSVGTGVQILAVSHTGIKLLQMVKGSKEASRRLRVLCAYSFADILF 2807

Qy 289 VECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKNSGYVIALPSYITDNCSLL 348

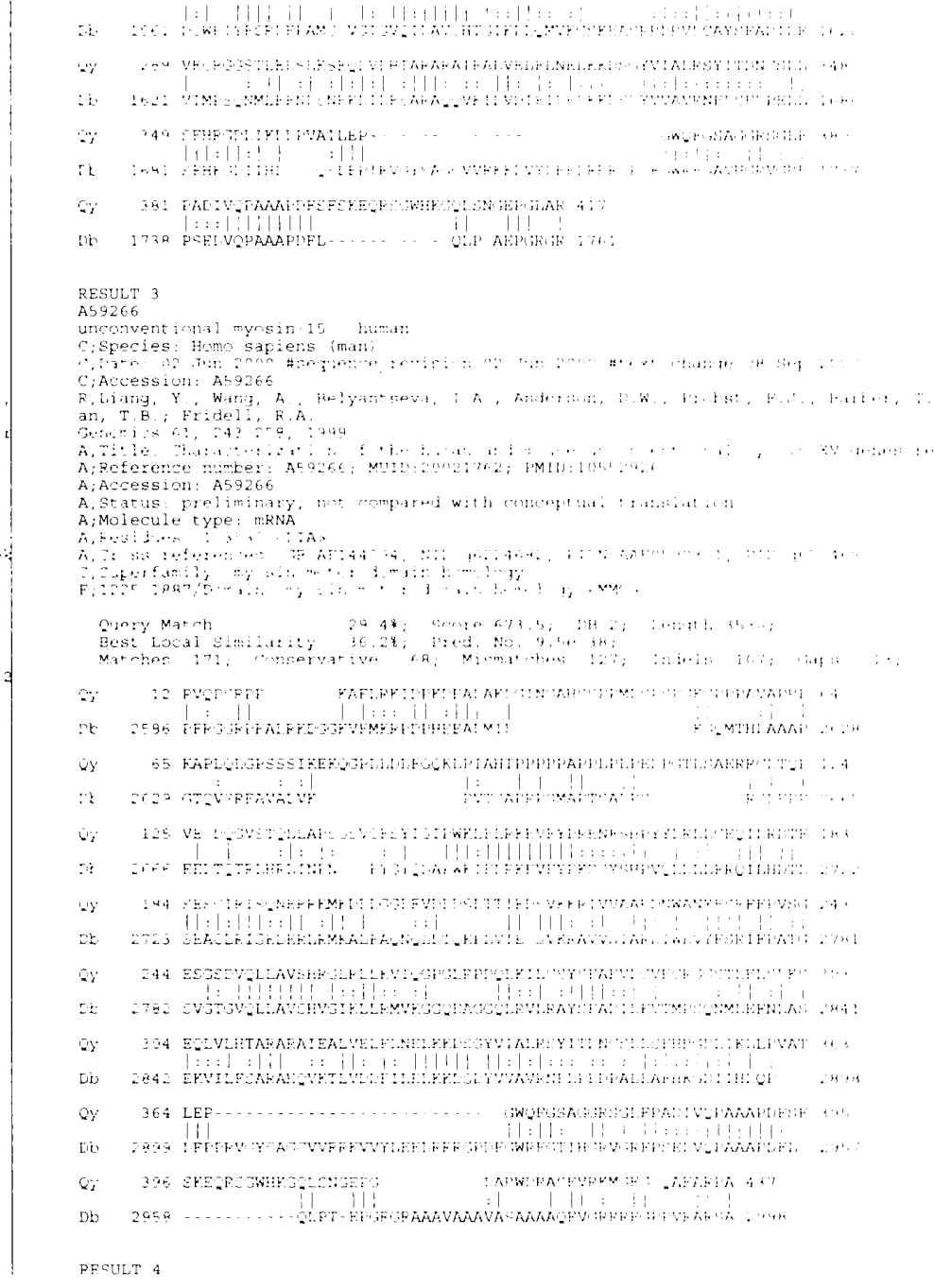


RESULT 2
T42386
unconventional myosin 15 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1999 #sequence_revision 02-Dec-1999 #text_change 08-Sep-2000
C;Accession: T42386; JE0268
R;Probst, F.J.; Fridell, R.A.; Raphael, Y.; Saunders, T.L.; Wang, A.; Liang, Y.; Morell, R.J.; Touchman, J.W.; Lyons, R.H.; Noben-Trauth, K.; Friedman, T.B.; Camper, S.A.
Science 280, 1444-1447, 1998
A;Title: Correction of deafness in shaker-2 mice by an unconventional myosin in a BAC transgene.
A;Reference number: JE0268; MUID:98269604; PMID:9603734
A;Accession: T42386
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1783 <PRO>
A;Cross-references: EMBL:AF057133; NID:g3176861; PID:AAC31.4.1
R;Wakabayashi, Y.; Takahashi, Y.; Fujiwara, Y.; Okano, H.; Mishina, Y.; Uchiki, T.; Yamamoto, M.
Biochem. Biophys. Res. Commun. 248, 666-669, 1998
A;Title: A novel type of myosin encoded by the mouse deafness gene shaker-2
A;Reference number: JE0268; MUID:98369604; PMID:9703981
A;Accession: JE0268
A;Molecule type: mRNA
A;Residues: 72-312,'KI',163-409,'R',411-654 <WAK>
A;Cross-references: GB:AB014811; NID:g4126464; PID:BAA33682.1; PID:BAA33683.1; PID:g4126465
A;Note: this sequence variant may cause mouse deafness
C;Genetics:
A;Gene: My; 15, shaker 2
A;Map position: 11
C;Function:
A;Description: may be required for actin organization in the hair cells of the cochlea
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;42-707/Domain: myosin motor domain homology <MMO>

Query Match 30.1%; Score 688; DB 2; Length 1783;
Best Local Similarity 34.0%; Pred. No. 3.4e-39;
Matches 174; Conservative 70; Mismatches 135; Indels 136; Gaps 13;

RESULT 3
A59266
unconventional myosin-15 - human
C;Species: Homo sapiens (man)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59266
R;Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; Miller, W.; Touchman, J.W.; Jones, M.; Lorenzo, I.; Oddoux, C.; Morell, R.J.; Ostrer, H.; Camper, S.A.; Bellman, S.; Kachar, B.; Fridell, R.A.; Friedman, T.B.; Friedl, R.A.
Genomics 61, 242-257, 1999
A;Title: Characterization of a functional myosin VI gene family.
A;Reference number: A59266; MUID:20021762; PMID:10512702
A;Accession: A59266
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1291 <LIA>
A;Cross-references: GB:AF144754; NID:g7144641; PID:AAF59154.1; PID:g7144642
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;1009-1087/Domain: myosin motor domain homology <MMO>

Query Match 29.4%; Score 673.5; DB 2; Length 1783;
Best Local Similarity 36.2%; Pred. No. 9.5e-38;
Matches 171; Conservative 68; Mismatches 127; Indels 107; Gaps 14;



RESULT 4

```
22298
ypothetical protein F46C3.3 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
;Accession: T22298
R;Costanzo, A
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19543
A;Accession: T22298
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2810 <WIL>
A;Cross-references: EMBL:Z66563, PIDN:CAA91469.1; GSPDB:GN00028; CESP:F46C3.3
A;Experimental source: clone F46C3
C;Genetics:
A;Gene: CESP:F46C3.3
A;Map position: X
A;Introns: 67/2; 142/3; 204/2; 310/3; 366/2; 421/2; 684/3; 739/3; 786/3; 823/2; 880/2; 1
1958/3; 1993/1; 2078/1; 2117/3; 2159/2; 2220/1; 2269/2; 2306/1; 2399/2; 2444/1; 2488/3;

  Query Match              8.9%;  Score 204.5;  DB 2;  Length 2810;
  Best Local Similarity   22.3%;  Pred. No. 1.4e-05;
  Matches   95;  Conservative   62;  Mismatches  176;  Indels   32,  Gaps   16;

Qy      19 NPAHSPPPMLSPSPGKGPPPAVAPRPKAPLQLGPSSSI---------------------- 76

Db    1876 NGNVVPPP---PSSYRAPSPAPTSGDRRGLNRLPQESYTEPAVKNTKGNNSEYLSHFFNL 1932

Qy      77 ----------KEKQGPLLDLFGQKL--PIAHTPPPPPAPPL--------PLPEDPGTLSA 116

Db    1933 DEQFAVIPHFKFTAKNAINMLSPDLRSPVPQPTEPPPPPPVREEWVIRDSVRPDPVTQGR 1992

Qy     117 EPPCL-TQPVEDGGVSTQLLAPSGSVCFSYTGTPWKLPLRKEVFYPRENFSHFYYLRLLC 175

Db    1993 IPGSFPPPPLVPQPPPPPVVAEKPAV--KFVKAPWKLTIRKEMFYPGEVLNDIQIIDQVF 2050

Qy     176 PQIIFCTFSESCTIKTSQNERPFMKDLLGGLEV-  DLDSLTTTEDSVKKRIVVAARDNWA 232

Db    2051 AQIVESCKKSYPYFIRQDPRKQVETVLRQYEIPPSDLNNQSNIHPDVKVAVIELAE LWF 2109

Qy     233 NYFSPPPPV----SGESGSDVQLLAVSHRGLRLLLKVTQGPGLRPDQLKILCSYSFAEVLG 288

Db    2110 LYPNQVYEVVSEKPPDESVSTI--FAISEHGIRLIVHT--PHDLENPLKIQDFPPPETIAD 2165

Qy     289 VPCPGGGTLPLSLKSEV-----  LVLHTAPAPAIEALVELPLNELKKDSGYVIALRSYI 341

Db    2166 VSLEANDILSVHVPRHEDEENAYSAVRIKTNQAPQIKKTLDPCL------SGGVVIPPPFV 2219

Qy     342 TPNQSLSFHPGDLIKLLPVATLEPGWQFGSAGGRSGLFPADIVQ-------PAAAPDPS 394

Db    2220 LARTGP---------ETPPVG----NWLYGKIENRFGFLLAQYVDSTGDMVPFIPHLTS 2266

Qy     395 FSKEQP 400

Db    2267 EDRDER 2272


RESULT 5
A59255
myosin VIIa, long form - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Feb-2001
C;Accession: A59255; S71932; S60263
R;Chen, Z.Y.; Hasson, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M
Genomics 36, 440-448, 1996
A;Title: Molecular cloning and domain structure of human myosin-VIIa, the gene product d
A;Reference number: A59255; MUID:97038686; PMID:8884267
A;Accession: A59255
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-2175 <CHE>
A;Cross-references: GB:U55208; NID:g1613787; PIDN:AAC50927.1; PID:g1613788
A;Experimental source: tissue type testis

R,Weil, D
submitted to the EMBL Data Library, November 1994
A;Reference number: S71932
A;Accession: S71932
A;Molecule type: mRNA
A;Residues: 96-334,'T',336-370,'N',372-388,'VTG',392-469,'P',471-563 <WEI>
A;Cross-references: EMBL:U17180; NID:g1235669
A;Note: the sequence is revised in GenBank entry HSU39226, release 114. (PIDN:AAB03679.1)
R;Weil, D.; Blanchard, S.; Kaplan, J.; Guilford, P.; Gibson, F.; Walsh, J.; Mburu, P.; M
Piet, D.; Munnich, A.; Steel, K.P.; Brown, S.D.M.; Petit, C.
Nature 374, 60-61, 1995
A;Title: Defective myosin VIIA gene responsible for Usher syndrome type 1B.
A;Reference number: S60263; MUID:95174880; PMID:7870171
A;Accession: S60263
A;Molecule type: mRNA
A;Residues: 96-157;199-255 <WEW>
A;Cross-references: EMBL:U17180; NID:g1235669
A;Note: the sequence is revised in GenBank entry HSU39226, release 114. (PIDN:AAB03679.1)
C;Comment: A alternate transcript from the same gene is shown in PIR:A59257.
C;Genetics:
A;Gene: MYO7a; USH1B
A;Map position: 11q13.5
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; ATP; nucleotide binding; P-loop
F;96-563/Domain: myosin motor domain homology (fragment) <MMOT>
F;158-165/Region: nucleotide-binding motif A (P-loop)

  Query Match              8.7%;  Score 199;  DB 2;  Length 2175;
  Best Local Similarity   24.1%;  Pred. No. 2.4e-05;
  Matches  108;  Conservative   49;  Mismatches  153;  Indels  140;  Gaps   18;

Qy      30 DFALAKLGINGAHSSPPM---LSPSPGKGP---PPAVAPRPKAPLQLGPSSSIKE---- 78

Db    1289 EEPIPFTPVSGTPTQPPSWLELQATKSKKPIMLFVTPMDGTTKTLLIDSATTAKEVNAL 1287

Qy      79 -KQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAEKSLTQEVPPQVSTQLLAP 137

Db    1288 ADKISLKDRPGFSLYIA---------------------- LPLPVSSLSPGSSDHVMD 1321

Qy     138 SGSVCFSYT---------GTPWKLPLRKEVFYPRENFSHF-YYLRLLCPQIIFCTFSESCT 188

Db    1322 AISQCEQYAKEQGAQERNAPWRLFFRKEVFIPWHSLSEDNVATNLIYEQVVPG------V 1375

Qy     189 PISQNERPFMKDLLG----------GLEVDLDSL---------------------------- 212

Db    1376 KPGEYRCEKEDDLAELASQQYFVDYGSEMILERLLNLVFTTIPDFEITFLKTLPFWAQLA 1435

Qy     213 ----------TTTEDSVKKRIVVAARDNWANYFSPPFPVSGFSG-- ---SLVSLLAVS 255

Db    1436 IAAHKKGIYAQRRTDAQFVKEDVVSYAPFKWPLLFSRFYEAYKFD-SLFFMCV IVAVN 1494

Qy     256 HPGLRLLKVTQGPGLRPDQLKILCSYSPAEVLGV-ECRGGSTLELS-    LKSFQLVIH 309

Db    1495 WTGVYFV---------DEQEQVLLELSFPEIMAVSSSRGAKITAPCFTLATKPGDEYTPT 1545

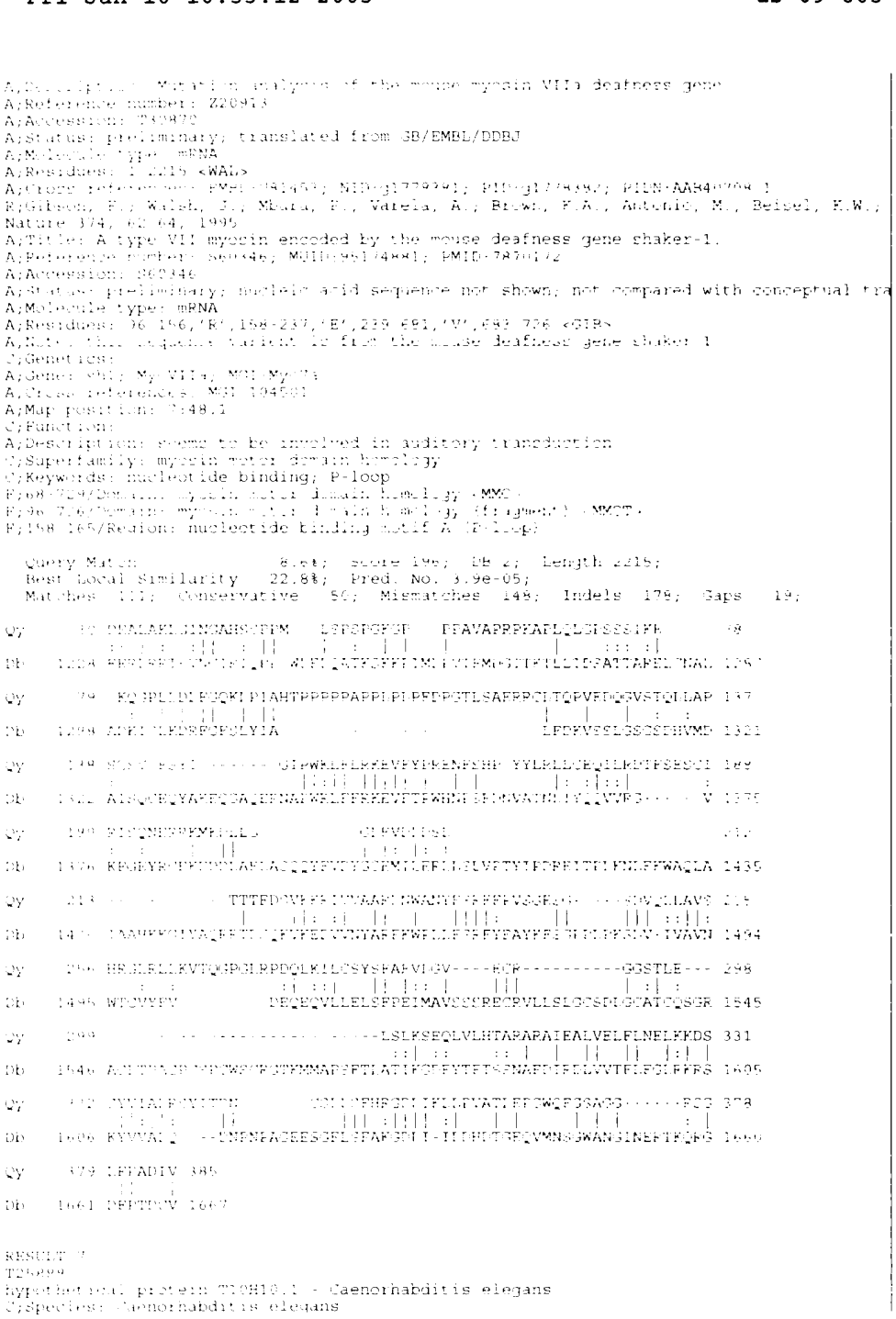
Qy     310 TAPAPAIEALVELPLNELKKDSGYVIALRSYITDN-------CSLLSFHPGDLIKLLPVA 362

Db    1546 SSNAEDIRDLVVTFLEGLPKPSKYVVALQ----DNFNIAGRLCFDLSFAPGDLI ILDHD 1600

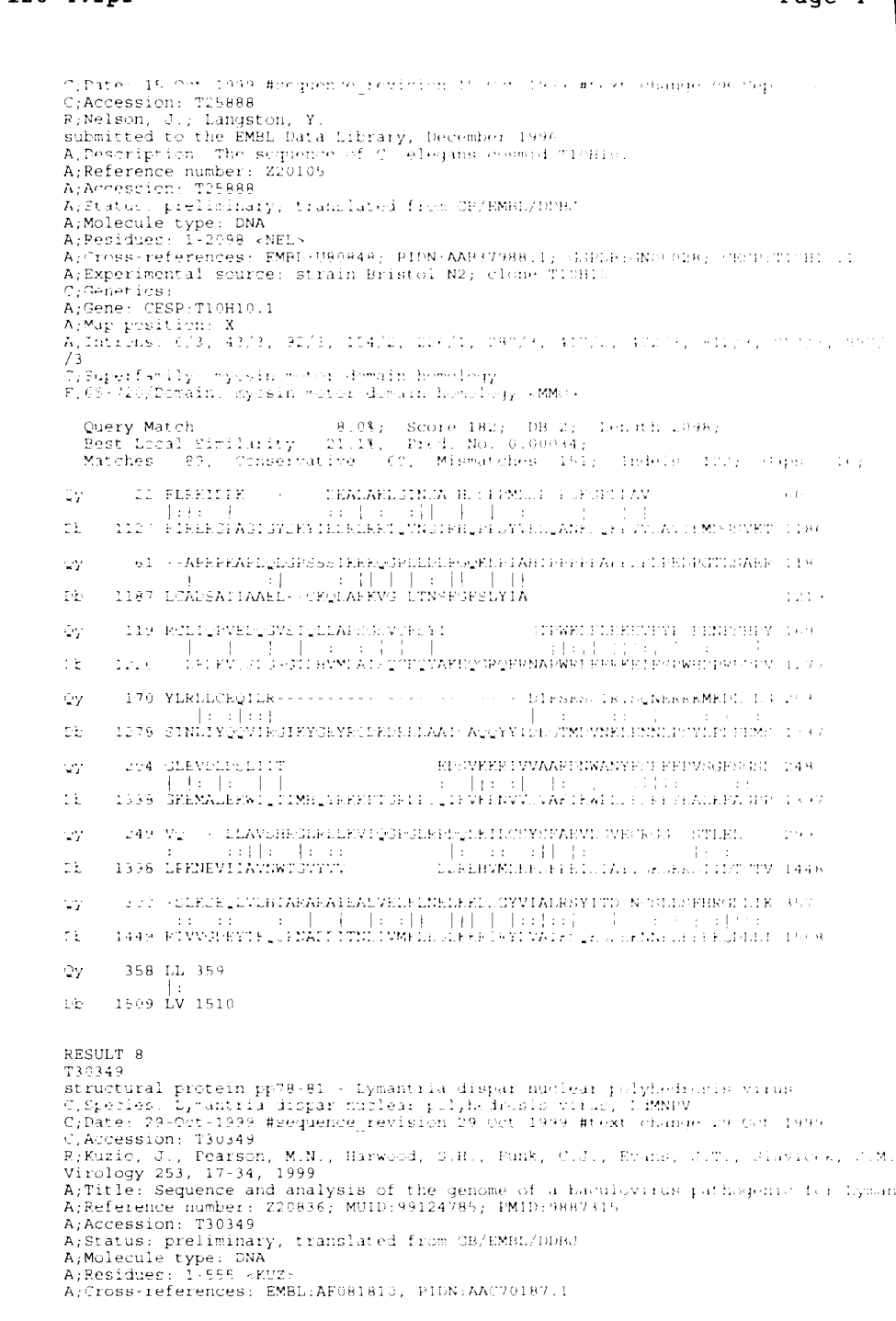
Qy     363 TLEPGWQFGSAGG------RSGLFPADIV 385

Db    1601 TGEQVMNSGWANGINERTKQRGDFPTDSV 1629


RESULT 6
T30870
myosin VIIa - mouse
N;Alternate names: gene shaker-1 protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Feb-2001
C;Accession: T30870; S60346
R;Walsh, J.; Mburu, P.; Gibson, F.; Liu, X.; Saw, D.; Steel, K.P.; Brown, S.D.M.
submitted to the EMBL Data Library, December 1996
```



RESULT 7
T25999
hypothetical protein T10H10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans



RESULT 8
T30349
structural protein pp78-81 - Lymantria dispar nuclear polyhedrosis virus
C;Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T30349
R;Kuzio, J., Pearson, M.N., Harwood, S.H., Funk, C.J., Evans, J.T., Slavicek, J.M.; E
Virology 253, 17-34, 1999
A;Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantri
A;Reference number: Z20836; MUID:99124786; PMID:9887315
A;Accession: T30349
A;Status: preliminary, translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-555 <KUZ>
A;Cross-references: EMBL:AF081810; PIDN:AAC70187.1

Query Match 6.5%; Score 148.5; DB 2; Length 555;
Best Local Similarity 22.7%; Pred. No. 0.012;
Matches 90; Conservative 45; Mismatches 139; Indels 123; Gaps 18;

Qy 5 RPGPVPV PVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPP--------MLSPSPGKG 55
Db 190 QPAPAPTRPIAPAPEPTRQETPIAPASEPARQESPIGSSAPEPIRQETPTGLFAPPP--P 247

Qy 56 PPPAVAPRPKAPLQLSPSSSIKPVQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDP-GTL 114
Db 248 PPPPPPPPPPEPLQ---------------QKSSAVPPPPPPPLPPPGAPLDPFGEI 288

Qy 115 SAEPRCLTQPVEDQGVSTQLLA---------PSGSVCFSYTGTPWKLF--LRKEVFY--- 160
Db 289 GQEVR PKPAAEPAPTDALPAEIRPGVQLKPATERAPTY--TPDALFAEIRQGVKLKPA 344

Qy 161 ---- -PRENFSHPYYLRLLCPQILRDTFSESCIRISQNERRFMKDLLGGLEVDLDSLT 213
Db 345 AERADEPPKPSSRAPLLL----------------EIENRDKIKLKKVAPPATEPPASAT 387

Qy 214 TTE--DSVKKRI----VVAARDNWANYFSPFFPVSGFSGSDVQLLAVSHRGLRLLKVTQG 267
Db 388 NNPLMQLINPFLESMYMSSAEESDANYTS-------SSWSDAEDDSLLRDALRIKLALLG 440

Qy 268 PGLPPDQLKILCSYSPAEVLGVECPGGSTLELSLKSEQLVLHTARARAIE---------- 317
Db 441 PRLSESERKRIAK RLAGAK----------LKSAEKTLDELQARAIEPDNPLLSPPY 486

Qy 318 -ALVELPLNELKKDSGYVIALRSYITDNCSLLSPHRG 353
Db 487 QLTAPLYLHDLKLFESAVLDL----------FNRG 511

RESULT 9
S16748
proline-rich protein - rape (fragment)
C;Species: Brassica napus (rape)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C;Accession: S16748
R;Roberts, M.R.; Foster, G.D.; Blundell, E.F.; Robinson, S.W.; Draper, J.; Scott, R.J.
submitted to the EMBL Data Library, August 1991
A;Description: Cloning and characterization of a proline-rich gene expressed specifically
A;Reference number: S16748
A;Accession: S16748
A;Molecule type: mRNA
A;Residues: 1-449 <ROB>
A;Cross-references: EMBL:X60376, NID:g42597, PID:g42597

Query Match 6.0%; Score 138; DB 2; Length 449;
Best Local Similarity 22.9%; Pred. No. 0.049;
Matches 83; Conservative 46; Mismatches 164; Indels 142; Gaps 20;

Qy 3 ARPGPVPVPVQ PSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKG------ 55
Db 33 QPKPPPAPTPSPCPPQPPKP-----QPK-------------PPPAPGPSPKPGPSPSPP 73

Qy 56 PPPAVAPRPKAPLQLSPSSSIKPVQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDP--- 111
Db 74 KPPPSPAPKPVPP----PSPS--PKPSP------PKPPAPSPKPSPPKPPAPSPPKPQPK 121

Qy 112 -----GTLSAERRCLTQPVE---DQGVSTQLLAPSGSVCFSYTGTPW 150
Db 122 TIPAVFFFGDSIFDTGNNNNLDTKLKCNYRPYGMDFPMGVATGKPS-NGRVASDYIS -- 177

Qy 151 KLFLPKEVFYPRENFSHPYYLRLLCEQILRDTFSESCIRISQNERRKMKDLLGGLE---- 206
Db 178 KYLGVKPIV--------PAYVDK----------KLQQNNELQQSDLLTGVSFASG 214

Qy 207 ------------YELPSLTTTEDSVKKR----------IVVAARDNW 231
Db 215 GAGYLPQTSESWKVTTMLDQLTYFQD-YKEFMKKLVGKKYTKKIVSKGAAIVVAGSNDLI 273

Qy 232 ANYFSR-----FFPVSGESGSDVQLLAVSHRGLRLLKVTQGP--GLRPDQ---L 275
Db 274 YTYFGNGAQHLKNDVDSFTTMMADSAASFVLQLYGY-GARRIGVIGTPPIGCTPSQRVKK 332

Qy 276 KILCS 280
Db 333 KKICN 337

RESULT 10
A59233
myosin VII-like protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: A59233
R;Anderson, J.B.; Yamashita, R.A.; Sellers, J.R.
submitted to GenBank, February 2000
A;Description: Complete cDNA for an unconventional Myosin (Class VII) in Drosophila mela
A;Reference number: A59233
A;Accession: A59233
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-2121 <AND>
A;Cross-references: GB:AF233269; PIDN:AAF34810.1
C;Genetics:
A;Gene: Myo28B1
A;Cross-references: FlyBase:FBgn0040299
A;Map position: 2, 28B1
C;Superfamily: myosin motor domain homology
F;69-725/Domain: myosin motor domain homology <MMO>

Query Match 6.0%; Score 137.5; DB 2; Length 2121;
Best Local Similarity 21.0%; Pred. No. 0.099;
Matches 100; Conservative 64; Mismatches 158; Indels 155; Gaps 21;

Qy 18 PPKAFLRKIDPKDEALAKLGINGAHSSPPM--LSPSPGKGP---PPAV----APRPKAPL 68
Db 1187 PPS---KEFEPHLRSFMKQGTAQLQATPSLQRLERTLVNGPPCQPPSLPELHAIRGRHPL 1243

Qy 69 QLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLTQPVEDQ 128
Db 1244 KLDIH-------- LMDGQQRRLQV---------- ... EAASTAREAVNQLTQCM 1276

Qy 129 GVSTQL----------LAPSG----------SVCFS-YTGTPWKLFLPKEVFYFRENFS 166
Db 1277 GLTDTFGFGLVMSLNGKLMPLGAGQEHVLDAISECEQSLTFAPWKLYIPKEMFATWYDPS 1336

Qy 167 -HPYYLRLLCEQILRDTFSESCIRISQNERRFMKDL----------LGGLEV------ 207
Db 1337 MDPKATQLIYKQILNG- ... LKCGEYPCPSEKDIAMVCALACFVEYGFGEILRLKPSE 1390

Qy 208 -------DL-------------------------------DSLTTTEDSVKKRIVVAARDN 230
Db 1391 ITAFVPSDLLAPGERAIENWSRLIAATTERSSYVFEQNPLLLEAQFPAPEDICLFAHLS 1450

Qy 231 WANYFSRFFPVSGESG----SDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSPAEV 286
Db 1451 WPMRHSRLFEVVRKEGPKLQSDELMLGINSAGLFLIDETE---------QVLASCCPSEV 1501

Qy 287 LGVECPGGSTLE-LSLKSEQLVLHTARARAIEALVELPLNELKKDSGYVIALRSYI --T 342
Db 1502 LKVHVESDDKLHVMTFQHVNPVLQCSSAQDANEVINYMLCNLPCFSGYGVAIDPVVPGDL 1561

Qy 343 DNCSLLSFHRGDLIKLLPVATLEPGWQF----------GSAGGRSGLFPADIVQFAA 389
Db 1562 EDCLVL--NPGDLIEFEAGVT---GAQLMAGNADCYPGCVNGQWGCFLAGNVRVLA 1613

RESULT 11
S21961
proline-rich protein APG - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C;Accession: S21961

R;Roberts, M.E.; Foster, G.D.; Blundell, P.P.; Robinson, S.W.; Draper, J.; Scott, R.J.
submitted to the EMBL Data Library, August 1991
A;Description: Cloning and characterization of a proline rich gene expressed specifically
A;Reference number: S16748
A;Accession: S21961
A;Molecule type: DNA
A;Residues: 1-534 <ROB>
A;Cross-references: EMBL:X60377; NID:g22598; PID:g22599
C;Genetics:
A;Gene: APG
A;Introns: 256/1; 299/3; 387/3; 470/1

Query Match 5.9%; Score 135.5; DB 2; Length 534;
Best Local Similarity 34.8%; Pred. No. 0.091;
Matches 40; Conservative 4; Mismatches 34; Indels 37; Gaps 6;

Qy 6 PSPVTVTVV 15
Db 73 PGPSPEPVAPPGPSPCPSPPPEP 113

Qy 69 AVAPPPAPLQLSPS 111
Db 114 PLPPPVPPPACPPTPPPPQPPPAP 159

RESULT 12
T37781
probable cytoskeleton assembly control protein - fission yeast (Schizosaccharomyces pomb
C;Species: Schizosaccharomyces pombe
C;Date: 01-Dec-1999 #sequence_revision 13-Dec-1999 #text_change 16-Dec-1999
C;Accession: T37781
R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21746
A;Accession: T37781
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1420 <OLI>
A;Cross-references: PMPI:P78714; EMBL:AB011381; PIDN:BAA31381.1; PID:g3702812
A;Experimental source: strain 972h ; cosmid c16E8
C;Genetics:
A;Gene: SPCC16E8.01
A;Map position: 1
A;Introns: 76/3

Query Match 5.9%; Score 135; DB 2; Length 1420;
Best Local Similarity 30.8%; Pred. No. 0.35;
Matches 107; Conservative 69; Mismatches 206; Indels 134; Gaps 23;

Qy 14 PKAPLRKIQ 22
Db 173 PHTTYLSMSACSAIASQRPGAGSVDAFTVSAFHSMVSFFPSFQPPSAAAFATSLFSDYNF 232

Qy 72 PSPSIKRPQQPPIP 85
Db 233 PPPPPPIPAVEQAALANEPIETYSSSFAVSFEIFFTTTPRQADFLPAFPPFPPPFTLPPQ 292

Qy 107 LPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFS 144
Db 293 STNTTQLTMTSRNVNNLSQQVNIFTPTATPSQFFRTT NASTRSTGTSSSMAHS 346

Qy 145 PSPEVFYIREMPEMP TYLRLLCE 168
Db 347 YTSTT PTLIFSLAYSLFMQHLPLAIICHIIFAVSSSSSMSNFAYEKWEVREV 406

Qy 169 LEVDLPGLTTTEDSVFKRIVVAARTNWANVFCP FFPVSGFST CDV 249
Db 407 VKFFFFPTPILAINMFSTVLTFTFTMTAAQVWFVTTIVNYSSSFFKHVFIFFNSDSGITSL 466

Qy 250 QLLAVSH 257
Db 467 HLHAPTTNSNNIIRALSVASAFAAIIFFIAAASSFMFKL PSFCALHFLNAASSAA 526

Qy 298 GVECPGGSTIEISIPSPQLVIHIAPAPAIPAIVPIPLNplpp 341
Db 476 GV NSSFTSSYEMST VVGSPSAPAHS RPIPLSSAVQPM TTV LYIPPI 525

Qy 342 TDNCSLLSFHSCPLIKLLPVATLEPGWQPGSAGPSGLSPPAIIV 386
Db 526 AFAADPLTV PANMPVVIVNPTASSPWWPSVPPPVVPSNFIKPITEGD ARSPPT 576

Qy 402 GWHKQQLSNGEPGLAPWDPASEVPEMSGQAELARP 436
Db 577 SPKSQQGSS----LSPPASKHESKHKPDSKHKARP 607

RESULT 13
S11515
formin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 09-Nov-1999
C;Accession: S11515
R;Woychik, R.P.; Maas, R.L.; Zeller, R.; Vogt, T.F.; Leder, P.
Nature 346, 850-853, 1990
A;Title: 'Formins': proteins deduced from the alternative transcripts of the limb
A;Reference number: S11515; MUID:90303291; PMID:2392154
A;Accession: S11515
A;Molecule type: mRNA
A;Residues: 1-1468 <WOY>
A;Cross-references: EMBL:X53599, NID:g52877, PIDN:CAA37618.1; PID:g52878

Query Match 5.9%; Score 134; DB 2; Length 1468;
Best Local Similarity 30%; Pred. No. 0.43;
Matches 105; Conservative 51; Mismatches 148; Indels 214; Gaps 25;

Qy 6 PSPVTVTVVSPLPTTPAPPTPPPPPPAPEL 34
Db 670 PPTLPTTLIPTPTT LTTSLSLPTPAPTTTTVTVSP 711

Qy 35 PAPLQLSPSSIPEPQGPLLPLPGQRLPIAHTPPPPPAPEL 77
Db 712 PPTTTTPPTTVPPPPM PPPPPPPPPPPPPPPPPPPP 758

Qy 108 PEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWPLPLPPEVPYPPENPSH 167
Db 959 PPPPG---------LAPPPPPGLSPGLSSSSSQG- YPPKPIAIP 998

Qy 168 P-------YYLRL-- -- --- --- LCEQILRDT PSESCIR 199
Db 999 PSQPMKPLYWIKQINDKSQWAAPILWPSLRPPHIPPPPPPPPPPPTQPPPPIPPA 1058

Qy 190 ISQNEPPPKMKDLLGG IPVDL DSLTTTEDSVPKPIVVA 226
Db 1051 YPPPNPVPPIIPLLPPPPSQTVPIIPSPIHLPMPPPQQAIPTVLPVVPTPTIAAIYPNP 1110

Qy 227 ARDNWANYFGRPPPVSGEDSDPVQLLAVGHRGLRLLRVTQPGLRPLGKILGVSTPAEV 286
Db 1111 AQEDELTKIRPVYETSPKE--EDLKLLPPPPQPPLH -ELAQIPNPAPPAQTIPPAVPSP 1169

Qy 287 LGVECPGGST-----LELSLPSEQLVLHIAPAPAIEAIVPLPLNGL 327
Db 1166 -------GIISLHRRVEIVIPASPGLIHMKSVPDILALILAPGNYMNPSNPTNPQAPSYP 1218

Qy 328 -------- - PPDSG- --YVIA--IR-YITQNSSISPM- - 360
Db 1219 LEILPKLPDVKGPENGMNLVDYVPYYLPYY PVT KHHLQPAS PGPDLPSLYPHI 1274

Qy 363 TLEPGWQPG-----SAGGPSGLPP 481
Db 1274 AVHPGPKSGLELPQEAGTLPSVPP 1297

RESULT 14
T05225
metosin b.p.-like FITS 162 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)

;Date: 23 Apr-1999 #sequence_revision 23 Apr-1999 #text_change 11-Jun-1999
;Accession: T05225
;Bevan, M.; Vitale, D.; Liguori, R.; Argiriou, A., De Simone, V.; Hoheisel, J.; Mewes,
ubmitted to the Protein Sequence Database, July 1998
;Reference number: Z15404
;Accession: T05225
;Molecule type: DNA
;Residues: 1-699 <BEV>
;Cross-references: EMBL:AL031032
;Experimental source: cultivar Columbia; BAC clone F17I5
;Genetics:
;Map position: 4
;Note: F17I5.160

Query Match 5.8%; Score 133; DB 2; Length 699;
Best Local Similarity 33.9%; Pred. No. 0.19;
Matches 37; Conservative 9; Mismatches 31; Indels 32; Gaps 6;

```
Qy       5 RPGPVPV-PVQPSRFPKAFLRKIPPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV--A 61
           :| |||  |||  :|||    :  || |::           |||   |:|   ||| |
Db     482 KPSPVPGRPVQKPQPPKESPQPDDPYDQSPVT-----KRRSPP---PAPVNSPPPPVYSP 533

Qy      62 PKPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPP--APPLPLP 108
           | |  |:   |                 |  |:|||||  :|| | |
Db     534 PPPPPPVHSPP-----------------PPVHSPPPPPVYSPPPPPP 563
```

RESULT 15
S24407
formin isoform IV - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 05 Nov-1999
C;Accession: S24407
R;Jackson-Grusby, L.; Kuo, A.; Leder, P.
Genes Dev. 6, 29-37, 1992
A;Title: A variant limb deformity transcript expressed in the embryonic mouse limb defin
A;Reference number: S24407; MUID:92112033; PMID:1339380
A;Accession: S24407
A;Molecule type: mRNA
A;Residues: 1-1206 <JAC>
A;Cross-references: EMBL:X62379; NID:g51552; PIDN:CAA44244.1; PID:g51553

Query Match 5.8%; Score 132; DB 2; Length 1206;
Best Local Similarity 20.3%; Pred. No. 0.45;
Matches 87; Conservative 49; Mismatches 125; Indels 168; Gaps 18;

```
Qy       6 PGPVPVPVQPSRFPKAFLRKIPPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAP-RP 64
           | |:| |:  |  ||                  || | | |   | | | |  |
Db     650 PPPLPPPLIPPPPP----------------------LPPGLGPLPPAPPIPPVCPVSP 685

Qy      65 KAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPL----------------PL 107
           |   | : :    ||               |||||  |||                   |
Db     686 PPPPPPPPPTPVPPSDGP------------PPPPPPPPPLPNVLALPNSGGPPPPPPPP 732

Qy     108 PEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSH 167
           |  ||          |    |:|  | : |                          |||:
Db     733 PPPPG--------LAPPPPPGLSFGLSSSSSQ------------------YPRKPAIE 764

Qy     168 P-------YYLRL----------------LCEQILRDT------FSESCIR--------- 189
           |       |: |:                | |  :|||      ||:   :
Db     765 PSCPMKPLYWTRIQINDKSQDAAPTLWDSLEEPHIRDTSEFEYLFSKDTTQQKKKPLSEA 824

Qy     190 -ISQNERRKMKDLLGG-----------LEVDL----DSLTTTEDSVKKRIVVA------ 226
            :|: :|:  || |               | :::     :: | :|||     :|
Db     825 YEKKNKVKKIIKLLDGKPSQTVGILISSLHLEMKDIQQAIFTVDDSVVDLETLAALYENR 884

Qy     227 ARDNWANYFSRFFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEV 286
           |:::       :::  | |   |::||    : |   .. | |           |.    |:|
Db     885 AQEDELTKIRKYYETSKE--EDLKLLDKPEQFLH--ELAQIPNFAERAQCIIFRAVFSE- 939

Qy     287 LGVECRGGST---  -LFISLKSFQLVLHTARARAIEALVELFLNELK------KDSGYV 334
                 | |      :|:  :: .  .||    : | ||:  | | :       :  ||
Db     940 -------GITSLHRKVEIVTRASKGLLHMKSVKDILALILAFGNYMNGGNKTPGQAEGYS 992
```

Qy 335 IALRSYITD 343
: : : |
Db 993 LEILPKLKD 1001

Search completed: January 10, 2003, 02:24:27
Job time : 26.534 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein protein search, using sw model

Run on: January 9, 2003, 20:02:30 ; Search time 12.632 Seconds
(without alignments)
1434.857 Million cell updates/sec

Title: US-09-803-126-4
Perfect score: 2287
Sequence: 1 MYQSRPGPVPVPVQPSRPPK.....WDRASEVRKMGEGQAEARPA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	689.5	30.1	3511	1	MY15_MOUSE	Q9qzz4 mus musculu
2	673.5	29.4	3530	1	MY15_HUMAN	Q9ukn7 homo sapien
3	196	8.6	2215	1	MY7A_HUMAN	Q13402 homo sapien
4	196	8.6	2215	1	MY7A_MOUSE	P97479 mus musculu
5	143	6.3	559	1	WAS1_HUMAN	Q92558 homo sapien
6	140.5	6.1	282	1	ATF5_HUMAN	Q9y2d1 homo sapien
7	138	6.0	449	1	APG_BRANA	P40603 brassica na
8	134	5.9	1468	1	FMN1_MOUSE	Q05860 mus musculu
9	133	5.8	533	1	RXRB_HUMAN	P28702 homo sapien
10	132	5.8	1206	1	FM14_MOUSE	Q05869 mus musculu
11	129	5.6	1426	1	CUT2_MOUSE	P70298 mus musculu
12	126.5	5.5	534	1	APG_ARATH	P40632 arabidopsis
13	126	5.5	512	1	GAG_SMSAV	P03330 simian sarc
14	125.5	5.5	1955	1	PC15_HUMAN	Q96qu1 homo sapien
15	125	5.5	497	1	WAS2_HUMAN	Q9y6w5 homo sapien
16	124.5	5.4	626	1	DF47_HUMAN	Q92570 homo sapien
17	124.5	5.4	1446	1	ZAP3_MOUSE	Q9r017 mus musculu
18	123.5	5.4	440	1	G3PT_MOUSE	Q64467 mus musculu
19	123	5.4	505	1	WASL_BOVIN	Q95167 bos taurus
20	123	5.4	3421	1	TEGU_HSVEB	P28955 equine herp
21	122.5	5.4	224	1	YC91_NPVAC	P41479 autographa
22	122	5.3	543	1	VP61_NPVAC	Q03209 autographa
23	122	5.3	626	1	EXTN_TOBAC	P13983 nicotiana t
24	121.5	5.3	261	1	PRP2_MOUSE	P05142 mus musculu
25	121.5	5.3	485	1	SSGP_VOLCA	P21997 volvox cart
26	121.5	5.3	555	1	GH1_CHLPE	Q9fpq6 chlamydomon
27	121	5.3	426	1	EXLP_TOBAC	Q03211 nicotiana t
28	121	5.3	815	1	MK07_HUMAN	Q13164 homo sapien
29	120.5	5.3	505	1	WASL_HUMAN	O00401 homo sapien
30	120.5	5.3	536	1	GAG_MLVCB	P27460 cas-br-e mu
31	120.5	5.3	1505	1	CUT2_HUMAN	Q14624 homo sapien
32	120	5.2	503	1	WAIP_HUMAN	O43516 homo sapien
33	120	5.2	806	1	MK07_MOUSE	Q9wvs8 mus musculu
34	119.5	5.2	501	1	WASL_RAT	O08816 rattus norv
35	119.5	5.2	1953	1	BNI1_YEAST	P41832 saccharomyc
36	119	5.2	538	1	GAG_MLVMO	P03332 moloney mur
37	118.5	5.2	817	1	VRP1_YEAST	P37370 saccharomyc
38	118.5	5.2	1567	1	FMN2_MOUSE	Q9jl04 mus musculu
39	118	5.2	1248	1	DIA1_HUMAN	O60610 homo sapien
40	117	5.1	316	1	CDNC_HUMAN	P42898 homo sapien
41	117	5.1	468	1	GAG_MSVMT	P32594 moloney mur
42	117	5.1	520	1	WASF_MOUSE	P70315 mus musculu
43	117	5.1	538	1	GAG_MSVMO	P03334 moloney mur
44	117	5.1	827	1	MOF_DROME	O02153 drosophila
45	117	5.1	961	1	FGD1_HUMAN	P96174 homo sapien

ALIGNMENTS

RESULT 1
MY15_MOUSE STANDARD; PRT; 3511 AA.
ID MY15_MOUSE
AC Q9QZZ4; O70395; Q9QWL6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin XV (Unconventional myosin-15).
GN MYO15A OR MYO15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20021762; PubMed=10552926;
RA Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
RA Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
RA Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,
RA Fridell R.A.;
RT "Characterization of the human and mouse unconventional myosin XV
RT genes responsible for hereditary deafness DFNB3 and shaker 2.";
RL Genomics 61:243-258(1999).
RN [2]
RP SEQUENCE OF 1168-2970 FROM N.A., AND VARIANT SH2 TYR-1779.
RC TISSUE=Embryo;
RX MEDLINE=98267310; PubMed=9603735;
RA Probst F.J., Fridell R.A., Raphael Y., Saunders T.L., Wang A.,
RA Liang Y., Morell R.J., Touchman J.W., Lyons R.H., Noben-Trauth K.,
RA Friedman T.B., Camper S.A.;
RT "Correction of deafness in shaker-2 mice by an unconventional myosin
RT in a BAC transgene.";
RL Science 280:1444-1447(1998).
RN [3]
RP SEQUENCE OF 1237-1823 FROM N.A., AND VARIANT SH2 TYR-1779.
RC STRAIN=C57BL/6;
RX MEDLINE=98369604; PubMed=9703981;
RA Wakabayashi Y., Takahashi Y., Kikkawa Y., Okano H., Mishima Y.,
RA Ushiki T., Yonekawa H., Kominami R.;
RT "A novel type of myosin encoded by the mouse deafness gene shaker-2.";
RL Biochem. Biophys. Res. Commun. 248:655-659(1998).
CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANEOUS
CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: IN THE DEVELOPING INNER EAR, EXPRESSED IN
CC COCHLEA AND VESTIBULAR APPARATUS. EXPRESSION APPEARS TO BE
CC RESTRICTED TO COCHLEAR NEUROSENSORY CELLS AND UPPER EPITHELIAL
CC LAYER OF MACULA SACCULA. ALSO EXPRESSED IN MACULA UTRICULI AND
CC CRISTAE AMPULLARIS OF THE SEMICIRCULAR CANALS. IN ADULT COCHLEAR
CC HAIR CELLS, HIGHEST EXPRESSION IN STEREOCILIA AND APICAL BODY.
CC -!- DISEASE: DEFECTS IN MYO15A ARE THE CAUSE OF SHAKER 2 (SH2), A
CC CONDITION CAUSING DEAFNESS, TIPTOEING BEHAVIOR, HEADTOSSING AND

```
CC       HYPERACTIVITY. AUDITORY HAIR CELLS OF AFFECTED ANIMALS HAVE VERY
CC       SHORT STEREOCILIA AND A LONG ACTIN CONTAINING PROTRUSION AT THEIR
CC       BASAL END.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1 LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF144095; AAF05904.1; -.
DR   EMBL; AF053130; AAC40124.1; -.
DR   EMBL; AB014510; BAA36562.1; -.
DR   HSSP; P13587; 1BR2.
DR   MGD; MGI:1261811; Myo15.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS00660; BAND_41_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_41_2; FALSE_NEG.
DR   PROSITE; PS50057; BAND_41_3; FALSE_NEG.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Disease mutation; Deafness.
FT   DOMAIN        1   1871       HEAD OR MOTOR DOMAIN.
FT   DOMAIN     1872   2013       NECK OR REGULATORY DOMAIN.
FT   DOMAIN     2014   3511       TAIL.
FT   DOMAIN     [illegible]   [illegible]       COILED COIL (POTENTIAL).
FT   DOMAIN     1776   1783       ACTIN BINDING (POTENTIAL).
FT   DOMAIN     [illegible]   1908       IQ 1.
FT   DOMAIN     [illegible]   1938       IQ 2.
FT   DOMAIN     [illegible]   2924       SH3.
FT   DOMAIN     [illegible]   3424       BAND 4.1 LIKE.
FT   NP_BIND    [illegible]   [illegible]       ATP (POTENTIAL).
FT   VARIANT    1779   1779       C -> Y (IN SH2).
FT   CONFLICT   [illegible]   1331       MISSING (IN REF. 2).
FT   CONFLICT   1579   1579       L -> R (IN REF. 3).
FT   CONFLICT   1955   1972       MISSING (IN REF. 2).
FT   CONFLICT   2077   2077       L -> M (IN REF. 2).
FT   CONFLICT   2139   2139       L -> F (IN REF. 2).
FT   CONFLICT   [illegible]   2353       V -> A (IN REF. 2).
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible]F98A[illegible]345B [illegible] CRC64;

  Query Match             30.1%;  Score 689.5;  DB 1;  Length 3511;
  Best Local Similarity   [illegible];  Pred. No. [illegible];
  Matches  194;  Conservative   72;  Mismatches  145;  Indels  147;  Gaps   15;

Qy          1 [illegible]AVAIF 63
Db       [illegible] [illegible]PRIVVS               [illegible]SVAFA [illegible]

Qy         64 PKAP-                 ------LQLGPS                  -CSIKE-KQGP-- 82
Db       2511 PKIPSKPVAVPILA[illegible]PWA [illegible]

Qy         83                          [illegible]SAPS [illegible]
Db       2571 GHPEAPPTD[illegible]HEEALMIL[illegible]LAVV[illegible]AVAMVKPVT SAPP [illegible]

Qy        113 P-IL              T[illegible]FPPENFSHH [illegible]
Db       2631 [illegible]FEYPTKEDNYSHH [illegible]

Qy        163 YYL[illegible]VVAAK [illegible]
Db       2689 V[illegible]MAK 2747

Qy        [illegible] [illegible]TARVIAS 288
Db       2748 DSWE[illegible]AYSFADILF [illegible]

Qy        289 VF[illegible]YVIALRSYITDN[illegible] 348
Db       2808 VTMP[illegible]YVVAVRNFLCFDPELL [illegible]

Qy        349 SFHRGDLIKLLPVATLEP---------                 -GWQFGSAGGRSGLF 38[illegible]
Db       2868 SFHRGDIIHL    [illegible]

Qy        381 PADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPG --          LARWDRASEVREMGEG 43[illegible]
Db       2925 PSELVQPAAAPDFL                   QLP AEPGRGRAAAVVAAVA[illegible]AAAAQEVGRRREG [illegible]

Qy        431 -QAEARPA 437
Db       2972 PPVRARSA 2979

RESULT 2
MY15_HUMAN
ID   MY15_HUMAN     STANDARD;      PRT;  3530 AA.
AC   Q9UKN7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin XV (Unconventional myosin-15).
GN   MYO15A OR MYO15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX   MEDLINE=20021762; PubMed=10552926;
RA   Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
RA   Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
RA   Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,
RA   Fridell R.A.;
RT   "Characterization of the human and mouse unconventional myosin XV
RT   genes responsible for hereditary deafness DFNB3 and shaker 2.";
RL   Genomics 61:243-258(1999).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A., AND VARIANTS DFNB3 TYR [illegible] AND PHE [illegible].
RX   MEDLINE=98267311; PubMed=9603736;
RA   Wang A., Liang Y., Fridell R.A., Probst F.J., Wilcox E.R.,
RA   Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K., Camper S.A.,
RA   Friedman T.B.;
RT   "Association of unconventional myosin MYO15 mutations with human
RT   nonsyndromic deafness DFNB3.";
RL   Science 280:1447-1451(1998).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. BY
CC       SIMILARITY, MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
CC       THE ACTIN RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY. ALSO EXPRESSED
```

```
CC       AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,
CC       PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN. IN THE
CC       PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.
CC   -!- DISEASE: Defects in MYO15A are the cause of an autosomal recessive
CC       form of nonsyndromic deafness (DFNB3).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licenseisib-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF144094; AAF05983.1; -.
DR   EMBL; AF051976; -. NOT_ANNOTATED_CDS
DR   HSSP; P10587; 1BR2.
DR   Genew; HGNC:7594; MYO15A.
DR   MIM; 602666; -.
DR   MIM; 600316; -.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00373; myosin_head; 1.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000165; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00660; BAND_4I_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_4I_2; FALSE_NEG.
DR   PROSITE; PS50057; BAND_4I_3; 1.
DR   PROSITE; PS50096; IQ; 3.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil, Repeat; SH3 domain;
KW   Calmodulin-binding; Disease mutation; Deafness.
FT   DOMAIN          1   1687       HEAD OR MOTOR DOMAIN.
FT   DOMAIN       1988   2027       NECK OR REGULATORY DOMAIN.
FT   DOMAIN       2030   3530       TAIL.
FT   DOMAIN       1323   1360       COILED COIL (POTENTIAL).
FT   DOMAIN       1792   1799       ACTIN-BINDING (POTENTIAL).
FT   DOMAIN       1902   1924       IQ 1.
FT   DOMAIN       1925   1954       IQ 2.
FT   DOMAIN       1955   1976       IQ 3.
FT   DOMAIN       2867   2953       SH3.
FT   DOMAIN       3206   3443       BAND 4.1 LIKE.
FT   DOMAIN       1315   1322       ATP (POTENTIAL).
FT   NP_BIND      1315   1322       N -> Y (IN DFNB3, FAMILY FROM BENGKALA).
FT   VARIANT      2111   2111       /FTId=VAR_010303.
FT                                  I -> F (IN DFNB3; INDIAN FAMILY).
FT   VARIANT      2113   2113       /FTId=VAR_010304.
FT
SQ   SEQUENCE   3530 AA;  395171 MW;  3D103923D4BCRE4A CRC64;

  Query Match           29.4%;  Score 673.5;  DB 1;  Length 3530;
  Best Local Similarity   26.2%;  Pred. No. 5.7e-36;
  Matches  171;  Conservative   68;  Mismatches  127;  Indels  167;  Gaps   13;

Qy      12 PVQPSRPP--    - KAPLPPIPPKPPALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRP 64
           |: :||          | |... || :|||  |                  ||   | |
Db    2595 PERGGRPEALRKDGGKVPMKPPDPHEEALMIL---------  --  -PGQMTHLAAAP 2628

Qy      65 KAPLQLGPSSGIKEKQGPLIDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLIQP 124
           :    : :|                  ||:  | | ||  ||           | | |
Db    2629 GTQVGREAVALVK-------------PVTSAPRPSMAPTSALPS---------RSLEPP 2665

Qy     125 VE-DQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSHPYYLRLLCEQILRDTF 183
           |  |    .|. ||          |  |||.|||||||:::|||  | ||  ||| ||
Db    2666 EELTQTRLHRLINDN  -FVGYQQAPWKIFLRKEVFYPFDSTSHFVQLDLLFPQILHDTL 2722

Qy     184 SESCIRTSQNERRYMKDILGGLEVDLDSLTTTEDSVKFFTVVAASDNWANYFSKEFFVSG 243
           || :|||.|.|| .|| |  ..|          || |||: :|  ||| |  |||| || :|
Db    2723 SEACLFTSEDFLFMFALFACNQIFTQKFLVTE SVKRAVVSIASRTWEVYFSPIFPATG 2781

Qy     244 ESGSDVQLLAVSHRGLFLLKVTQGPGLKPDQLKTLCSYSEAEVLGVEGPGGSTLELSLKS 303
           |: |||||||| |.||:|. .|          || .| :||||:| |           : || :| |
Db    2782 SVGTGVQLLAVSHVGIKLLPMVKGGQEAGGQLRVLPAYSFADILFVIMFSQMMLEPNLAS 2841

Qy     304 EQLVLHTAPARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIKLLPVAT 363
           |...| .|||  .  ||. |: |||||| ||.|:|. . .. .||:|:||:| | |
Db    2842 EKVILFSARAHQVKTLVDDFILELKKDSDYVVAVRNFLTEDPALLAFHFGDIIHLQR--  2898

Qy     364 LEP----------------------------GWQFGSAGGRGGLFPADIVQPAAAPDFSF 395
           |||                            ||:||:  || | ||:::||||||||
Db    2899 LEPPRVGYSAGCVVRKKVVYLEELRPPGPDPGWRFGTIHGRVGRFPSELVQPAAAPDFL- 2957

Qy     396 SYEQPSGWHKGQLSNGEPG----------LARWDRASEVRKMGEG-QAEARPA 437
                    ||  |||                :|   | ||  :  ||    |||
Db    2958 -----------QLPT-EPGRGRAAAVAAAVASAAAAQEVGRRREGPPVRARSA 2998


RESULT 3
MY7A_HUMAN              STANDARD,       PRT;  2215 AA.
ID   MY7A_HUMAN     STANDARD,       PRT;  2215 AA.
AC   Q13402; Q92821; Q92822; Q13321; P78427; Q14785;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin VIIa.
GN   MYO7A OR USH1B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria, Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1, 3, 6 AND 7), AND DEVELOPMENTAL
RP   EXPRESSION
RC   TISSUE=Retina;
RX   MEDLINE=96194949; PubMed=8622919;
RA   Weil D., Levy G., Sahly I., Levi-Acobas F., Blanchard S.,
RA   El-Amraoui A., Crozet F., Philippe H., Abitbol M., Petit C.;
RT   "Human myosin VIIA responsible for the Usher 1B syndrome: a predicted
RT   membrane-associated motor protein expressed in developing sensory
RT   epithelia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3232-3237(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).
RC   TISSUE=Testis;
RX   MEDLINE=97038686; PubMed=8884267;
RA   Chen Z.-Y., Hasson T., Kelley P.M., Schwender B.J., Schwartz M.F.,
RA   Ramakrishnan M., Kimberling W.J., Mooseker M.S., Corey D.P.;
RT   "Molecular cloning and domain structure of human myosin-VIIa, the gene
RT   product defective in usher syndrome 1B.";
RL   Genomics 36:440-449(1996).
RN   [3]
RP   SEQUENCE OF 166-196 FROM N.A.
RC   TISSUE=Epithelium, Leukocyte, and Liver;
RX   MEDLINE=94294418; PubMed=8022818;
RA   Bement W.M., Hasson T., Wirth J.A., Cheney R.E., Mooseker M.S.;
RT   "Identification and overlapping expression of multiple unconventional
RT   myosin genes in vertebrate cell types.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6549-6553(1994).
RN   [4]
RP   SEQUENCE OF 1-1075 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=96003870; PubMed=7568024;
RA   Hasson T., Heintzelman M.B., Santos-Sacchi J., Corey D.P.,
RA   Mooseker M.S.;
```

```
RT   "Expression in cochlea and retina of myosin VIIA, the gene product
RT   defective in Usher syndrome type 1B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:4815-4819(1995).
RN   [5]
RP   SEQUENCE OF 90-564 FROM N.A., AND VARIANTS USH1B
RC   TISSUE=Retina;
RX   MEDLINE=95174880; PubMed=7870171;
RA   Weil D., Blanchard S., Kaplan J., Guilford P., Gibson F., Walsh J.,
RA   Mburu P., Varela A., Levilliers J., Weston M.D., Kelley P.M.,
RA   Kimberling W.J., Wagenaar M., Levi-Acobas F., Larget-Piet D.,
RA   Munnich A., Steel K.P., Brown S.D.M., Petit C.,
RT   "Defective myosin VIIA gene responsible for Usher syndrome type 1B.",
RL   Nature 374:60-61(1995).
RN   [6]
RP   SEQUENCE OF 79-678 FROM N.A.
RX   MEDLINE=97074487; PubMed=9070921;
RA   Kelley P.M., Weston M.D., Chen Z.Y., Orten D.J., Bass S.,
RA   Lowther G.E., Bond J., Talmadge C.B., Ing P., Mooseker M.S.,
RA   Corey D.P., Dumont J., Kimberling W.J.;
RT   "The genomic structure of the gene defective in Usher syndrome type 1B
RT   (MYO7A).";
RL   Genomics 40:73-79(1997).
RN   [7]
RP   DEVELOPMENTAL EXPRESSION
RX   MEDLINE=94440126; PubMed=8049737;
RA   El Amraoui A., Sahly I., Picaud S., Sahel J., Abitbol M., Petit C.;
RT   "Human Usher 1B/mouse shaker-1: the retinal phenotype discrepancy
RT   explained by the presence/absence of myosin VIIA in the photoreceptor
RT   cells.";
RL   Hum. Mol. Genet. 5:1171-1178(1996).
RN   [8]
RP   VARIANTS USH1B H 212, C 212, H 302, Q 468 INS, L 503 AND Q 490.
RX   MEDLINE=97356489; PubMed=9302236;
RA   Weston M.D., Kelley P.M., Overbeck L.D., Wagenaar M., Orten D.J.,
RA   Hasson T., Chen Z.Y., Corey D.P., Mooseker M.S., Sumegi J.,
RA   Gibson F., Walsh J., Mburu P., Varela A., Levilliers J.,
RA   Weston M.D., Kelley P.M., Overbeck L.D., Wagenaar M., Orten D.J.,
RA   Hasson T., Chen Z.Y., Corey D.P., Mooseker M.S., Sumegi J.,
RA   Gibson F., Walsh J., Mburu P., Varela A., Levilliers J.,
RA   Weston M.D., Kelley P.M., Overbeck L.D., Wagenaar M., Orten D.J.,
RA   Hasson T., Chen Z.Y., Corey D.P., Mooseker M.S., Sumegi J.,
RA   Gibson F., Walsh J., Mburu P., Varela A., Levilliers J.,
RA   Kimberling W.J.;
RT   "Myosin VIIA mutation screening in 189 Usher syndrome type 1
RT   patients.";
RL   Am. J. Hum. Genet. 59:1074-1083(1996).
RN   [9]
RP   VARIANTS USH1B ARG 24, ALA 324 AND THR 326, AND POLYMORPHISMS.
RX   MEDLINE=98014437; PubMed=9326942;
RA   Adato A., Weil D., Kalinski H., Pel Or Y., Ayadi H., Petit C.,
RA   Korostishevsky M., Bonne-Tamir B.;
RT   "Mutation-profile of all 49 exons of the human myosin VIIA gene, and
RT   haplotype analysis, in Usher 1B families from diverse origins.",
RL   Am. J. Hum. Genet. 61:813-821(1997).
RN   [10]
RP   VARIANTS USH1B ARG 24, SER 765 AND GLY 2132, AND POLYMORPHISMS.
RX   MEDLINE=98067678; PubMed=9367678;
RA   Levy G., Levi-Acobas F., Blanchard S., Gerber S., Larget-Piet D.,
RA   Chedal V., Liu X.-Z., Newton V., Steel K.P., Brown S.D.M., Munnich A.,
RA   Kaplan J., Petit C., Weil D.;
RT   "Myosin VIIA gene heterogeneity: is the main role played by frequent
RT   Usher syndrome type 1B?";
RL   Hum. Mol. Genet. 6:111-116(1997).
RN   [11]
RP   VARIANT DFNB2 PRO 244.
RX   MEDLINE=97419477; PubMed=9171832;
RA   Liu X.-Z., Walsh J., Mburu P., Kendrick-Jones J., Cope M.J.,
RA   Steel K.P., Brown S.D.M.,
RT   "Mutations in the myosin VIIA gene cause non-syndromic recessive
RT   deafness.";
RL   Nat. Genet. 16:188-190(1997).
RN   [12]
RP   VARIANT DFNB2 ILE 199.
RA   MEDLINE=97419476; PubMed=9171831;
RA   Weil D., Represa J., El-Amraoui A., Levy G., Levi-Acobas F., Dodu M.,
RA   Ayadi H., Petit C.;
RT   "The autosomal recessive isolated deafness, DFNB2, and the Usher 1B
RT   syndrome are allelic defects of the myosin-VIIA gene.",
RL   Nat. Genet. 16:191-193(1997).
RN   [13]
RP   VARIANT DFNA11 986 ALA  LYS 988 DEL.
RX   MEDLINE=98016404; PubMed=9354784;
RA   Liu X.-Z., Walsh J., Tamagawa Y., Kitamura K., Nishizawa M.,
RA   Steel K.P., Brown S.D.M.;
RT   "Autosomal dominant non-syndromic deafness caused by a mutation in the
RT   myosin VIIA gene.";
RL   Nat. Genet. 17:268-269(1997).
RN   [14]
RP   VARIANTS USH1B PRO 451 AND GLN 1602.
RX   MEDLINE=98386290; PubMed=9718356;
RA   Liu X.-Z., Hope C., Walsh J., Newton V., Ke X.-M., Liang C.-Y., Xu L.-R.,
RA   Zhou J.-M., Tamagawa Y., Steel K.P., Bradley D., Brown S.D.M.;
RT   "Mutations in the myosin VIIA gene cause a wide phenotypic spectrum,
RT   including atypical Usher syndrome.";
RL   Am. J. Hum. Genet. 63:909-912(1998).
RN   [15]
RP   VARIANT USH1B PRO 1067.
RX   MEDLINE=99244058; PubMed=10364543;
RA   Adato A., Kalinski H., Weil D., Chaib H., Korostishevsky M.,
RA   Bonne-Tamir B.;
RT   "Possible interaction between USH1B and USH3 gene products as implied
RT   by apparent digenic deafness inheritance.";
RL   Am. J. Hum. Genet. 65:261-265(1999).
RN   [16]
RP   VARIANTS USH1B, AND POLYMORPHISMS.
RX   MEDLINE=99140287; PubMed=10094549;
RA   Janecke A.R., Mains M., Sadeghi M., Grundmann K., Apfelstedt-Sylla E.,
RA   Zrenner E., Rosenberg T., Gal A.;
RT   "Twelve novel myosin VIIA mutations in 34 patients with Usher syndrome
RT   type I: confirmation of genetic heterogeneity.";
RL   Hum. Mutat. 13:133-140(1999).
RN   [17]
RP   VARIANTS USH1B LYS 1170 AND CYS 1719.
RX   MEDLINE=99274392; PubMed=10447383;
RA   Cuevas J.M., Espinos C., Millan J.M., Sanchez F., Trujillo M.J.,
RA   Ayuso C., Perez-Garrigues H., Najera C.;
RT   "Identification of three novel mutations in the MYO7A gene.";
RL   Hum. Mutat. 14:181-181(1999).
CC   -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO A FIXED FILAMENTOUS
CC       RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RHODOPSIN
CC       SYNAPTIC VESICLE COMPONENTS ARE REMOVED FROM THE OUTER SEGMENT BY
CC       DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE POLARITY OF STEREOCILIA
CC       DURING THE DYNAMIC MOVEMENTS OF THE BUNDLES. MAY PLAY A ROLE IN HAIR CELL
CC       VESICLE TRAFFICKING OF AMINOGLYCOSIDES, WHICH ARE KNOWN TO INDUCE
CC       OTOTOXICITY (BY SIMILARITY).
CC   -!- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE WITH ATPASE
CC       FORMATION OF A COILED-COIL ROD.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE). IN THE PHOTORECEPTOR
CC       CELLS, MAINLY LOCALIZED IN THE INNER AND OUTER SEGMENTS AS
CC       WELL AS IN THE SYNAPTIC ENDING REGION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS, 1, 2, 3, 4,
CC       4, 5, 6 AND 7, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE PIGMENT EPITHELIUM AND THE
CC       PHOTORECEPTOR CELLS OF THE RETINA. ALSO FOUND IN KIDNEY, LIVER,
CC       TESTIS, CHOREA, LYMPHOCYTE. NOT EXPRESSED IN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: DETECTED IN OPTIC CUP IN 5.5 WEEKS OLD
CC       EMBRYOS. EXPRESSED IN RETINAL PIGMENT EPITHELIUM, COCHLEAR AND
CC       VESTIBULAR NEUROEPITHELIA, AND OLFACTORY EPITHELIUM AT 8 WEEKS. AT
CC       19 WEEKS, PRESENT IN BOTH PIGMENT EPITHELIUM AND PHOTORECEPTOR
CC       CELLS. AT 24-28 WEEKS, EXPRESSION IN PIGMENT EPITHELIUM AND
CC       PHOTORECEPTOR CELLS INCREASES. PRESENT IN PIGMENT EPITHELIUM AND
CC       PHOTORECEPTOR CELLS IN ADULT.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF USHER SYNDROME TYPE 1B
CC       (USH1B), ALSO KNOWN AS USHER SYNDROME NON-AMERICAN VARIETY. USHER
CC       SYNDROME TYPE 1 IS AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY
CC       PROFOUND CONGENITAL SENSORINEURAL DEAFNESS, CONSTANT VESTIBULAR
CC       DYSFUNCTION AND PREPUBERTAL ONSET OF PROGRESSIVE RETINITIS
CC       PIGMENTOSA LEADING TO BLINDNESS. USHER SYNDROME IS THE MOST COMMON
CC       CAUSE OF COMBINED DEAFNESS AND BLINDNESS IN DEVELOPED COUNTRIES.
CC       USH1B APPEARS AS A PRIMARY CYTOSKELETAL PROTEIN DEFECT.
```

```
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB2 OR NSPD2) WITH
CC       VARIABLE VESTIBULAR DYSFUNCTION AND VARIABLE AGE OF ONSET.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA11) WITH ONSET
CC       AFTER COMPLETE SPEECH ACQUISITION AND SUBSEQUENT GRADUAL
CC       PROGRESSION.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 IQ DOMAINS.

  Query Match              8.6%;  Score 196,  DB 1;  Length 2215;
  Best Local Similarity   22.0%;  Pred. No. 5.1e-05;
  Matches  110;  Conservative   48;  Mismatches  151;  Indels  178,  Gaps   18;

Qy      30 DEALAKLGINGAHSSPPM---LSPSPGKGP---PPAVAPRPKAPLQLGPSSSIKE----- 78
           :| | :  :||  : ||     |  :  |     |        :   :: ||
Db    1279 EERLPPTPVNGTRTQPPSWLELQATKSKKPIMLPVTFMDGTTKTLLTDSATIAELCNAL 1457

Qy      79 -KQGPLLDLFGQKLPIANTPPPPPAPPLPLPEDPGTLSAERPCLTQPVEDQGVSTQLLAP 137
            :  | | ||  | ||                                |  |  |  :  .
Db    1458 ADKISLKDPPGFSLYIA----------------------------LFDKVSSLGSGSFHVMD 1321

Qy     138 PQSVCFSYT--    ----GTPWPLFIFYEVFYPPENFSHP-YYLRLMCEQILEDTPSESCI 188
           :  | | |          ||:|| | |||| |   |     |: .:.:|             :
Db    1320 ATSQTFQVAPFCGACFPNAPWPLFFKKEVFTPWHSFSEDNVATNLIYQQVVRG------V 1375

Db    1374 RFGEYRIFFRPDLAPHASQQYFVDYGSEMILFRLLNLVPTYIPDRELTPLKTLEKWAQLA 1433

Qy     213 ------------TTTEDSVKKPIVVAARDNWANVFSPFFPVSGESG-----SDVQLLAVS 265
                       |  ||:  :| |  ||  |   ||||:      ||     :|| : |||.
Db    1434 IAAHKKGIYAQRRTDAQFVFFDVVSYARPFWPLLFSRPYEAYKFSGPSLPKNDV-IVAVN 1492

Db    1493 HRGLRLLKVTQGPGLRPDQLKILCSYSPAEVLGV---------------------------- 1545

Qy     256 HRGLRLLKVTQGPGLRPDQLKILCSYSPAEVLGV---------------------------- 289
           |:  :          :| ::|    | ||  |                              |
Db    1495 WTGVYFV---------DEQEQVLLELSFPEIMAVSSSRECRVWLSLGCSDLGCAAPHSGW 1545

Qy     290 ------------ECRGGSTLELS-----LKSEQLVLHTARARAIEALVELFLNELKKDS 331
                       |||  |  |      .| ::    :: | |  | ||  ||  |:| |
Db    1546 AGLTPAGPCSPCWSCPGAKTTAPSFTLATIKGDEYTFTSSNAEDIRDLVVTFLEGLRYPS 1605

Qy     332 GYVIALRSYITDN-  -----CSLLSFHRGDLIKLLPVATLEPGWQFGSAGG------RSG 378
           ||:|||    ||          ||| .|||| :|   | |      | |  |      : |
Db    1606 KYVVALQ-----PNFNPAGEESGFLSFAKGDLI-ILDHDTGEQVMNSGWANGINEFTKQPG 1659

Qy     379 LFPADIV 385
           | | | |
Db    1660 LFPTDCV 1667


RESULT 4
MY7A_MOUSE
ID   MY7A_MOUSE     STANDARD;      PRT;  2215 AA.
AC   P97479;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin VIIa.
GN   MY07A OR MYO7
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95174881; PubMed=7870172;
RA   Gibson F., Walsh J., Mburu P., Varela A., Brown K.A., Antonio M.,
RA   Beisel K.W., Steel K.P., Brown S.D.M.;
RT   "A type VII myosin encoded by the mouse deafness gene shaker-1.";
RL   Nature 374:62-64(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98343888; PubMed=9680294;
RA   Mburu P., Liu X.-Z., Walsh J., Saw D. Jr., Cope M.J., Gibson F.,
RA   Kendrick-Jones J., Steel K.P., Brown S.D.M.;
RT   "Mutation analysis of the mouse myosin VIIA deafness gene.";
RL   Genes Funct. 1:131-203(1997).
CC   -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN
CC       RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF PHOTORECEPTORS
CC       SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTORS
CC       DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE RIGIDITY OF STEREOCILIA
CC       DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE (BY SIMILARITY).
CC       INVOLVED IN HAIR-CELL VESICLE TRAFFICKING OF AMINOGLYCOSIDES,
CC       WHICH ARE KNOWN TO INDUCE OTOTOXICITY
CC   -!- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE THROUGH THE
CC       FORMATION OF A COILED-COIL ROD (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE RETINA, COCHLEA, KIDNEY AND
CC       LIVER.
CC   -!- DEVELOPMENTAL STAGE: IN THE INNER EAR OF THE 1- 5 DAY OLD EMBRYO,
CC       EXPRESSED ONLY IN THE COCHLEAR AND VESTIBULAR SENSORY HAIR CELLS.
CC       IN ADDITION, EXPRESSION ALSO OCCURS IN THE EPITHELIAL CELLS OF THE
CC       SMALL INTESTINE, HEPATOCYTES, AND CHOROIDAL PLEXUS.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF THE SHAKER-1 (SH-1)
CC       PHENOTYPE WHICH AFFECTS ONLY THE INNER EAR. SH-1 HOMOZYGOTE
CC       MUTANTS SHOW HYPERACTIVITY, HEAD-TOSSING AND CIRCLING DUE TO
CC       VESTIBULAR DYSFUNCTION, TOGETHER WITH TYPICAL NEUROEPITHELIAL TYPE
CC       COCHLEAR DEFECTS INVOLVING DYSFUNCTION AND REGRESSIVE
CC       DEGENERATION OF THE ORGAN OF CORTI
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 BAND 4.1-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U81453; AAB40708.1; -
DR   HSSP; P08799; 1MND.
DR   MGD; MGI:104519; Myo7a
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000066; SH3; 1.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00660; BAND_41_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_41_2; FALSE_NEG.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS50096; IQ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Calmodulin-binding; Disease mutation; Deafness.
FT   DOMAIN        1    729       MYOSIN HEAD-LIKE.
FT   DOMAIN      745    765       IQ 1.
FT   DOMAIN      768    788       IQ 2.
```

```
FT   DOMAIN      791    811       IQ 3.
FT   DOMAIN      814    834       IQ 4.
FT   DOMAIN      837    857       IQ 5.
FT   DOMAIN      858    935       COILED COIL (PROBABLE).
FT   DOMAIN     1255   1441       BAND 4.1 LIKE 1.
FT   DOMAIN     1603   1672       SH3.
FT   DOMAIN     1899   2139       BAND 4.1-LIKE 2.
FT   NP_BIND     158    165       ATP (PROBABLE).
FT   DOMAIN      622    639       ACTIN-BINDING (PROBABLE).
FT   VARIANT     241    241       R -> P (IN SH-1).
FT   VARIANT     502    502       R -> P (IN SH-1).
SQ   SEQUENCE   2215 AA;  254847 MW;  34740DD1941AB2F0 CRC64;

  Query Match             8.6%;  Score 196;  DB 1;  Length 2215;
  Best Local Similarity   22.8%;  Pred. No. 5.1e-05;
  Matches  111;  Conservative   50;  Mismatches  148;  Indels  178;  Gaps   19;

Qy         30 DEALAKLGINGAHSCPPM   ICPSPGKGP   PPAVAPPPKAPLQLGPSSSIKE       78
Db       1228 EEPLPPTPVNGTRTQPPCWLRLQAIPSPPFIMIPVTPMPITIPTLLTI-ATTAPEL-NAL 1287

Qy         79  KQGPLLDLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAP 137
Db       1288 ADPI----------LTIA-----------------------LPDKVSSLGSSGLHVMD 1321

Qy        138 SGSVCFSYT          GTPWRIFLPKEVFYPPENFCHP YYLRLLC EQILRPTFSFSCI 188
Db       1322 AISQCPQYAFPQGAQEKNAIWPLPPPKFVPTPWHNPGPDNVATNLIYQLVVRG        V 1375

Qy        189 RISQNERKKMKDLLG          GLEVDLDSL                         212
Db       1376 KPGEYRCEKKEDDLAELASQQYFVDYGSEMILERLLSLVPTYIPDREITPLKNLEKWAQLA 1435

Qy        213            - -TTILIGVRRKIVVAAKNWANYFGPFPPVGGESG     SEVQLLAVG 255
Db       1436 IAAEKKGIYAQEKTLLLKVKEDVVNYAPPKWPLLPSRPYEAYKFSGPPLPKSDV IVAVN 1494

Qy        256 HPGLPRLPKVTQGPGLPPPQLKILCTYSPAEVLGV- --PTP   --- - GGSTLE-    298
Db       1495 WTGVYFV-          -DEQEQVLLELSFPEIMAVSSSREGRVLLSLGCSDLGCATCQSGR 1545

Qy        299              ----- -------LSLKSEQLVLHTARAPAIEALVELFLNELKKDS 331
Db       1546 A[illegible]HMMAPSPTLAT[illegible]SNAE[illegible]FDLVVTFLEGLPKRS 1605

Qy        332 DTVIALRSYITDN          CGLLSFHRGDLIKLLPVATLEPGWQFGSAGG      RGG 378
Db       1606 KYVVALQ    DNPNPAGEPGGFLSFAKGDLI ILDHPTGPQVMNSGWANGINERTKQRG 1660

Qy        379 LFPADIV 385
Db       1661 CFPTDCV 1667


Result 6
WAS1_HUMAN
ID   WAS1_HUMAN     STANDARD;      PRT;   559 AA.
AC   Q92558;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Wiskott-Aldrich syndrome protein family member 1 (WASP family
DE   protein member 1) (Verprolin homology domain-containing protein 1).
GN   WASF1 OR SCAR1 OR KIAA0269.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=97191644; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98402540; PubMed=9732292;
RA   Bear J.E., Rawls J.F., Saxe C.L. III;
RT   "SCAR, a WASP-related protein, isolated as a suppressor of receptor
RT   defects in late Dictyostelium development.";
RL   J. Cell Biol. 142:1325-1335(1998).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99108169; PubMed=9889097;
RA   Machesky L.M., Insall R.H.;
RT   "Scar1 and the related Wiskott-Aldrich syndrome protein, WASP,
RT   regulate the actin cytoskeleton through the Arp2/3 complex.";
RL   Curr. Biol. 8:1347-1356(1998).
CC   -!- FUNCTION: DOWNSTREAM EFFECTOR MOLECULE INVOLVED IN THE
CC       TRANSMISSION OF SIGNALS FROM TYROSINE KINASE RECEPTORS AND SMALL
CC       GTPASES TO THE ACTIN CYTOSKELETON.
CC   -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX.
CC   -!- SUBCELLULAR LOCATION: DOT-LIKE PATTERN IN THE CYTOPLASM.
CC       CONCENTRATED IN RAC REGULATED MEMBRANE RUFFLING AREAS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN. LOWLY EXPRESSED IN
CC       TESTIS, OVARY, COLON, KIDNEY, PANCREAS, THYMUS, SMALL INTESTINE,
CC       AND PERIPHERAL BLOOD.
CC   -!- DOMAIN: BINDS ARP2/3 COMPLEX THROUGH THE C-TERMINAL REGION AND
CC       ACTIN THROUGH VERPROLIN HOMOLOGY (VPH) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 WH2 DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D87459; BAA13399.1; -.
DR   EMBL; AF134303; AAD33052.1; -.
DR   Genew; HGNC:12732; WASF1.
DR   MIM; 605035; -.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
DR   SMART; SM00246; WH2; 1.
KW   Actin-binding.
FT   DOMAIN      278    283       POLY-PRO.
FT   DOMAIN      322    332       POLY-PRO.
FT   DOMAIN      348    359       POLY-PRO.
FT   DOMAIN      369    374       POLY-PRO.
FT   DOMAIN      424    435       POLY-PRO.
FT   DOMAIN      497    514       WH2.
SQ   SEQUENCE   559 AA;  61652 MW;  44B4527BDB77BC6E CRC64;

  Query Match             6.3%;  Score 143;  DB 1;  Length 559;
  Best Local Similarity   24.5%;  Pred. No. 0.027;
  Matches   67;  Conservative   36;  Mismatches   97;  Indels   74;  Gaps   [illegible];

Qy          1 MYQGHP      HVPVPVIPGPIIPAPIPPIIPI ALAP[illegible]   PPMLSPSPG 53
Db        303 LIENRPQSPATGRTPVFVSPTLPPPP     PPLPSALSTSSLRASMTSTPPPPVPPPP  [illegible]

Qy         54 KGPPPAVA------PPPKAPLQLGP-      -    ---SSSIKKKQGPLLDLPGQKLPIAHT  96
Db        357  -PPPATALQAPAVPPPPAPLQIAPGVLHPAPPPIAPPLVQPSPPV     ARAAPVCETV 410

Qy         97 -- --          PPPPPAPPLP     LPEDPGTLSAERRCLTQPVEDQGV-STQLLAPS 138
Db        411 PVHPLPQGEVQGLPPPPPPPPLPPPGIRPSSPVTVTA   -LAHP  PPGLHPTPSTAPG 464

Qy        139 SGVCFSYTGTPWRIFLPKEVFYPPENFCHPYYLRLLCE  QILRP[illegible]IP[illegible]QNERR 196
```

```
Db    465 PHVPLMPPSPP------SQVIFASEPKPHPSTLPVISDARSVLLEAIRKGIQLRKVEEQR 518

Qy    197 KMKDLLGGLEVDLDSLTTTEDSVKKRIVVAARDN 230

Db    519 EQEAKHERIENDVATI--    -LSPPIAVEYSDS 546

RESULT 6
ATF5_HUMAN
ID   ATF5_HUMAN     STANDARD;       PRT;   282 AA.
AC   Q9Y2D1, Q9NRQ3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cyclic-AMP-dependent transcription factor ATF-5 (Activating
DE   transcription factor 5) (Transcription factor ATFx).
GN   ATF5 OR ATFX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20558615; PubMed=11087824;
RA   White J.H., McIllhinney R.A.J., Wise A., Ciruela F., Chan W.Y.,
RA   Emson P.C., Billinton A., Marshall F.H.;
RT   "The GABAB receptor interacts directly with the related transcription
RT   factors CREB2 and ATFx.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:13967-13972(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kohroki J., Tanaka K.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 161-282 FROM N.A.
RX   MEDLINE=99303793; PubMed=10373550;
RA   Pati D., Meistrich M.L., Plon S.E.;
RT   "Human Cdc34 and Rad6B ubiquitin-conjugating enzymes target repressors
RT   of cyclic AMP-induced transcription for proteolysis.";
RL   Mol. Cell. Biol. 19:5001-5013(1999).
CC   -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
CC       (CONSENSUS: 5'GTGACGT(A/C)(A/G)-3'), A SEQUENCE PRESENT IN MANY
CC       VIRAL AND CELLULAR PROMOTERS.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER (POTENTIAL). INTERACTS WITH GABAB
CC       RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a licence agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF305687; AAG22558.1; -.
DR   EMBL; AB021663; BAA78477.2; -.
DR   EMBL; AF101388; AAD28370.1; -.
DR   TRANSFAC; T04877; -.
DR   Genew; HGNC:790; ATF5.
DR   MIM; 606398; -.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW   Multigene family.
FT   DOMAIN      123    139       POLY-PRO.
FT   DOMAIN      186    194       POLY-PRO.
FT   DNA_BIND    212    230       BASIC MOTIF.
FT   DOMAIN      236    250       LEUCINE ZIPPER (PROBABLE).
FT   CONFLICT    161    163       LLA -> RHE (IN REF. 3).
SQ   SEQUENCE   282 AA;  30674 MW;  DDB2F907CA0215A0 CRC64;

  Query Match              6.1%;  Score 140.5;  DB 1;  Length 282;
  Best Local Similarity   27.1%;  Pred. No. 0.017;
  Matches   65;  Conservative   18;  Mismatches   72;  Indels   85;  Gaps   12;

Qy      6 PGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAH-------SSPPMLSPSPGKGPPP 58

Db     85 PGTLP---QPSPTP--------PDLEAMASLLKKELEQMEDFFLDAPPLPPPSPPPLPPP 133

Qy     59 AVAPRPKAPLQLGPSSSIKEKQGPLLDLF-----------GQKLPIAHTP-PPPPAPPLP 106

Db    134 PLPPAPSLPLSL-PSFDL--PQPPVLDTLDLLAIYCRNEAGQE-EVGMPPLPPPQQPPPP 189

Qy    107 LPEEPGTLSAEPPCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLREEVFYPRENFS 166

Db    190 SPPQPSP--------------------LAP---------------------YP----- 201

Qy    167 HPYYLRLLCEQILRDTFSESCIRISQNERRKMKDLLG---GLEVDLDSLTTTEDSVKKRI 223

Db    202 HPATTRGDPKQKKRDQNEEAALPYPQPPPAEGEALEGECQGLEARNRELKERAESVEREI 261

RESULT 7
APG_BRANA
ID   APG_BRANA      STANDARD;      PRT;   449 AA.
AC   P40603;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Anter-specific proline-rich protein APG (Protein CEX) (Fragment).
GN   APG OR CEX.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94004980; PubMed=8401599;
RA   Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA   Draper J., Scott R.;
RT   "Gametophytic and sporophytic expression of an anther-specific
RT   Arabidopsis thaliana gene.";
RL   Plant J. 3:111-120(1993).
CC   -!- TISSUE SPECIFICITY: FOUND IN ANTHER, ONLY IN MALE FERTILE PLANTS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC       MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC       MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC       MATURATION.
CC   -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X60376; CAA42924.1; -.
DR   PIR; S16748; S16748.
DR   InterPro; IPR001087; Lipase_GDSL.
DR   Pfam; PF00657; Lipase_GDSL; 1.
DR   PROSITE; PS01098; LIPASE_GDSL_SER; 1.
FT   NON_TER       1      1
FT   ACT_SITE    132    132       BY SIMILARITY.
FT   ACT_SITE    428    428       POTENTIAL.
SQ   SEQUENCE   449 AA;  48779 MW;  9EFB6A3AB28EEA15 CRC64;

  Query Match              6.0%;  Score 138;  DB 1;  Length 449;
  Best Local Similarity   22.7%;  Pred. No. 0.044;
```

```
  Matches   93;  Conservative   26;  Mismatches  109;  Indels  147;  Gaps   14;

Qy        1 QDPPVPVPVPVPVQ PSRPPPAPLPPIEPPEPALAKLSINGAHSSPPMLSPPPPKPPPAVAP  60
            | :: : | |   | :|||       ||                ||   |||  |
Db       82 QPKPPPAPTPSPCPPQPPPF  ---QPP--------------PPPAPGPSPPPPSPSPPP  73

Qy       61  PPPAVAPPPKAPLQLSPSSSIKEPLSLLPLPSQPLEIAHTIPPPAPPLPLPELP    111
             .||| |||  |    || |   | .       | |     | || ||| | ||  |
Db       74 P---APPVIP    IPPS  PPIP     - PPPPAPGPPPPPPPPAPGPPPPQNP 121

Qy      112         ------ ------GTLSREPPSLTQPPVP   PQSVSPLSLAISSGVVPSYTPHW 163
                            .  : :|  :|       ||:|   : :| | |   |
Db      122 TIPAVIPPPIIIPTTPPPPPLPTPIP TVPPIS IKEPMGVAP PPP NGPVAPVTIS  - 170

Qy      164 KLPLPKKPVPVPPTNPTHPVVLPLLTPQILPTTPSPSCIPIPQNPPPPMKPLLQCLE    206
                ||:          |:|                 :: ||    :  ||| |:
Db      171 PTLPTPEIP         IATVPA             ---  PLQLPNNELLIPPLLLTPVCPASG 214

Qy      207                   VDLDSLTTPEDSVKKP                    IVVAARPNW 231
                              :| ||   :|  :||                    :|  : |
Db      215 GASYLPQTSESWPEVTIMLPQLPIPQP VPPPMPPIVPPPTPPIVSPSAAAIVVARSNFLI 274

Qy      232 ANVFNF            PPPVQGEQSSDVQLLAVSHPGLPLLKVTQQP  GLRPPQ---L 275
            ||   :            |  :  :| |    |    : | | :  |    |   | |
Db      275 YTTPGNIAQHLPHTVIPPTTMNAI AAAPVTIQVTV TAPPISVIGTIIIPPTTISIPVPP 332

Qy      276 PIEDS 280
            : :|:
Db      333 KKIPN 337


RESULT 8
FMN1_MOUSE
ID   FMN1_MOUSE      STANDARD;      PRT;  1408 AA.
AC   Q09962;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Formin 1 (Protein I/II/III) (Limb deformity protein).
GN   FMN OR LD.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney, and Testis;
RX   MEDLINE=93263291; PubMed=2392150;
RA   Woychik R.P., Maas R.L., Zeller R., Vogt T.F., Leder P.;
RT   "'Formin'-like proteins deduced from the alternative transcripts of the
RT   limb deformity gene.";
RL   Nature 346:850-853(1990).
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=99124469; PubMed=9119367;
RA   Wang C.-C., Chan D.C., Leder P.;
RT   "The mouse formin (Fmn) gene: genomic structure, novel exons, and
RT   genomic mapping.";
RL   Genomics 39:303-311(1997).
RN   [3]
RP   PHOSPHORYLATION.
RX   MEDLINE=94[illegible]; PubMed=8[illegible];
RA   Vogt T.F., Jackson Grusby L., Rush J., Leder P.;
RT   "Characterization of phosphoproteins isoforms encoded by the mouse limb deformity
RT   locus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5554-5558(1993).
CC   -!- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
CC       AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
CC       INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; IA (SHOWN HERE), IB,
CC       II, III AND IV (AC [illegible]), ARE PRODUCED BY ALTERNATIVE SPLICING.
CC       A VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
CC       DIFFERENT SITE (ANALOGOUS, EXIST WITHIN ALL OF THESE.
CC   -!- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
CC       LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
CC       GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
CC   -!- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
CC       IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESEN'HYMAL
CC       COMPARTMENT, PREDOMINANTLY IN THE [illegible].
CC   -!- DISEASE: DISRUPTION OF THE FMN GENE CAUSES LIMB AND KIDNEY
CC       KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
CC       THE EPITHELIAL COMPARTMENT OF THE LB NEPHROS AND MESONEPHROS.
CC   -!- PTM: PHOSPHORYLATED ON SERINE AND POSSIBLY THREONINE RESIDUES.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FH1) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. APPEARING
CC       SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute for Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licence agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53599; CAA37668.1;  -.
DR   PIR; S11515; S11515.
DR   MGD; MGI:101815; Fmn.
DR   InterPro; IPR003104; FH2.
DR   InterPro; IPR001265; Formin.
DR   Pfam; PF02181; FH2; 1.
DR   PRINTS; PR00828; FORMIN.
DR   SMART; SM00498; FH2; 1.
KW   Nuclear protein; Developmental protein; Alternative splicing;
KW   Phosphorylation; Coiled coil.
FT   DOMAIN      723    792       COILED COIL (POTENTIAL).
FT   DOMAIN      870    970       FH1 (PRO-RICH).
FT   DOMAIN      985   1426       FH2.
FT   DOMAIN     1305   1378       COILED COIL (POTENTIAL).
FT   DOMAIN      198    203       POLY-SER.
FT   DOMAIN      861    864       POLY-SER.
FT   DOMAIN      885    892       POLY-PRO.
FT   DOMAIN      911    925       POLY-PRO.
FT   DOMAIN      929    940       POLY-PRO.
FT   DOMAIN      951    962       POLY-PRO.
FT   DOMAIN      966    970       POLY-PRO.
FT   DOMAIN      977    981       POLY-SER.
FT   VARSPLIC   1252   1287       MISSING (IN ISOFORM IB).
FT   VARSPLIC    625    722       MISSING (IN ISOFORM II).
FT   VARSPLIC    626    627       IA -> SV (IN ISOFORM II).
FT   VARSPLIC    629   1469       MISSING (IN ISOFORM III).
SQ   SEQUENCE   1408 AA,  [illegible] MW,  [illegible] CRC64;

  Query Match             5.9%;  Score 134;  DB 1;  Length 1408;
  Best Local Similarity   20.8%;  Pred. No. 0.35;
  Matches  105;  Conservative   51;  Mismatches  148;  Indels  310;  Gaps   17;

Qy        6 PSPVPVPVQPSRPPPAPLPEPIEPPEPALAKLSINGAHSSPPMLSPPPPKPPPAVAP 64
            | |:| |: |  ||                          :| |  | :  : |  | |
Db      876 PPPLPPPLIPPPPPP--------------           --LPPGPGPLPTAPPIPPVCPVSP 911

Qy       65 PAPLQLSPSSSIPEPVGPLLDLPGQPLPIAHTPPPPPAPPL                  PL 107
              |    | : :    ||              |||||  |||                  |
Db      912 PPPPPPPPPTPVPPSPMP  --------   PPPPPPPPPPINVIALPNGSPPPPPPPP 958

Qy      108 PEPEPSPLAPSPTLPVPPVPVPLPIAP [illegible] PLPLVPPPNPQH 157
            |  ||           |    :|  | : |                  :||:
Db      959 PPPPG---------LAPPPPPGPSPGPSSSSQ - -          - -  YPPKPAIE 990

Qy      158 P       YPLAP              LPLQPLRNDT      PPPSPIP          189
            |       |: |:              | |   :|||      ||:  :
Db      991 PSQPMKPLYWTPIQINDRSQDAAPTLWPSFPPHIPDTSPPPYIPSPITDLPKKPLSPA 1050

Qy      190 -LQLNEPPPMPIIPSP- - -       LEVEL    IIIPTTPTPPPPVVA       226
             :|: :|:  || |               | :::     :: | :||,     :|
```

```
Db   1091 YFFYNKVVKIIKLLDGKRSQTVGILISSLHLEMKDIQQAIFTVDDSVVDLETLAALYENR 1110

Qy    227 AFDNWANYPSFFFPVSGFSGSDVPLLAVSHFGLFLLVVTQGPGLRPDQLKILCSYSFAEV 286

Db   1111 AQEPELTKIPKYYYFTSKF--EDLKLLDKPEQFLH--ELAQIPNFAERAQCTIFPAVFSE- 1165

Qy    287 LGVECPGGST----LFLSLKSEQLVLHTARARAIEALVELFLNEL-------------- 327

Db   1166 -------GITSLHPFVKFIVTRASKGLLHMKSVKPILALILAFGNYMNGGNPTRGQADGYS 1218

Qy    328 ------ --- KKDSG-----YVIA--LRSYITDNCSLLSFH------RG-DLIKLLPVA 362

Db   1219 LEILPKLKDVASRDTNSMNLVDYYVPVYYLPYY--DQC---KHHDQEASCPGKDLFSLYFHI 1273

Qy    363 TLEPGWQFG--    SAGGPSGLFP 381

Db   1274 AVHPQRKSGLELKQEAGTDKSVFP 1297


RESULT 9
RXRB_HUMAN                  STANDARD;      PRT;   533 AA.
ID   RXRB_HUMAN
AC   P28702, P29703;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Retinoic acid receptor RXR-beta.
GN   RXRB OR NR2B2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=92255...; PubMed=1315968;
RA   Fleischhauer K., Park J.H., Disanto J.P., Marks M.S., Ozato K.,
RA   Yang S.Y.;
RT   "Isolation of a full-length cDNA clone encoding a N-terminally
RT   variant form of the human retinoid X receptor beta.";
RL   Nucleic Acids Res. 20:1801-1801(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=92127595; PubMed=1310259;
RA   Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,
RA   Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S.,
RA   Chambon P.;
RT   "Purification, cloning, and RXR identity of the HeLa cell factor with
RT   which RAR or TR heterodimerizes to bind target sequences
RT   efficiently.";
RL   Cell 68:377-395(1992).
RN   [3]
RP   REVISIONS
RX   MEDLINE=93045602; PubMed=1330328;
RA   Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,
RA   Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S.,
RA   Chambon P.;
RL   Cell 71:887-887(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Numasawa T., Koga H., Ueyama K., Maeda S., Sakou T., Harada S.,
RA   Leppert M., Inoue I.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Corella A., Vergara A., Paez G., de Miguel C., Encio I.;
RT   "Molecular cloning and characterization of the human HEXFB gene and 5'
RT   flanking region.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Tubby B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
CC       9-CIS RETINOIC ACID (9C-RA).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; a long form (shown
CC       here) and a short form, are produced by alternative splicing.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS. A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63522; CAA45087.1; -.
DR   EMBL; M84820; AAA60293.1; -.
DR   EMBL; AF065396; AAC18599.1; -.
DR   EMBL; AF120161; AAD13794.1; -.
DR   EMBL; AL031228; CAA20239.1; -.
DR   EMBL; BC001167; AAH01167.1; -.
DR   PIR; S22647; S22647
DR   PIR; S39753; S39753.
DR   HSSP; P19793; 2NLL.
DR   TRANSFAC; T01334; -.
DR   Genew; HGNC:10478; RXRB
DR   MIM; 180246; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA binding; Nuclear protein;
KW   Zinc-finger; Multigene family; Alternative splicing.
FT   DOMAIN        1    204       MODULATING (BY SIMILARITY).
FT   DNA_BIND    205    270       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     205    225       C4-TYPE.
FT   ZN_FING     241    265       C4-TYPE.
FT   DOMAIN      271    330       HINGE.
FT   DOMAIN      331    533       LIGAND-BINDING (BY SIMILARITY).
FT   DOMAIN       61    198       PRO-RICH.
FT   CONFLICT    112    112       S -> T (IN REF. 2).
SQ   SEQUENCE   533 AA;  56321 MW;  D00B4FE93AC16A04 CRC64;

  Query Match              5.8%;  Score 133;  DB 1,  Length 533;
  Best Local Similarity   22.6%;  Pred. No. 0.12;
  Matches   93;  Conservative   46;  Mismatches  153;  Indels  120;  Gaps   19;

Qy       4 SRPGPVPVPVQPSRPPKAPLPKPIDPKPDEALAKLGINGAHSSPPMLSP---------- 50

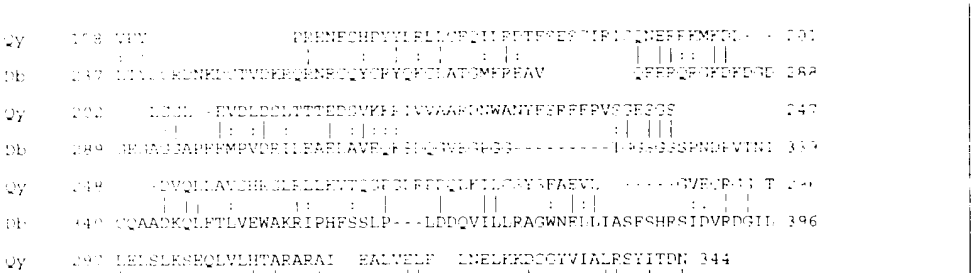
Db      86 SSPNPLPQGVPPPSPPGPPL----PPSTA-PSLGGSGAPPPPPMPPPPLGSPPPVIGSDM 140

Qy      51 -SPGKGPP-------PAVAPRPKAPLQL-----GFSSSIKEVQGPLLDLFGQKLPIAHTP 97

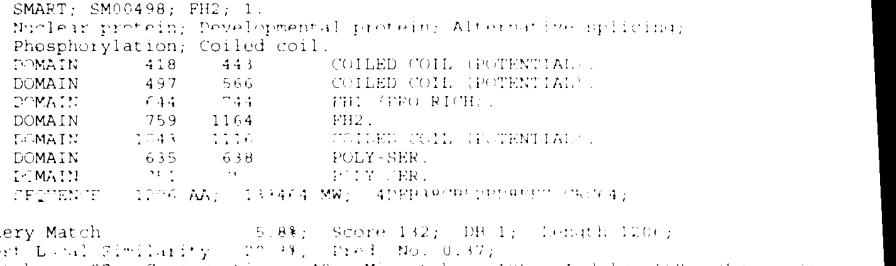
Db     141 GSPGLPPPAPPGFSGPVSSFQINSTVSLPGGGSGPPEDVKF-  PVLGVPG-----LHCP 192

Qy      98 PPPPAPPLPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLPLPFE 157

Db     193 PPPGGPG--------AGKRLCAICGDRSSGKHTGVTGDFGPGP----PKRTIRKD 236
```



RESULT 10
FMI4_MOUSE
ID FMI4_MOUSE STANDARD; PRT; 1200 AA.
AC Q9Z0S9;
DT 01-JUN-1999 (Rel. 29, Created)
DT 01-JUN-1999 (Rel. 29, Last sequence update)
DT 15-MAY-2000 (Rel. 39, Last annotation update)
DE Formin 1 isoform IV (Limb deformity protein).
GN FMN OR LD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94110033; PubMed=1333280;
RA Grusby Jackson L., Kuo A., Leder P.;
RT "A variant limb deformity transcript expressed in the embryonic mouse
RT limb defines a novel formin.";
RL Genes Dev. 6:29-37(1992).
CC -!- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB AND MAY HAVE A
CC FUNCTION IN DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING
CC SPECIFIC DIFFERENTIATED STATES.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; IA (AC Q9Z0S9-1), IB (AC
CC Q05860-1), II (AC Q05860), III (AC Q05860) AND IV (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING. A VARIATION IN SPLICING IS SEEN
CC AMONG DIFFERENT TISSUES AND DIFFERENT SIZE TRANSCRIPTS EXIST
CC WITHIN ANY ONE TISSUE.
CC -!- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
CC HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.
CC -!- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL
CC ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING
CC LIMB BUD.
CC -!- PTM: PHOSPHORYLATED ON SERINE AND POSSIBLY THREONINE RESIDUES
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FH1) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. CAPPUCCINO
CC SUBFAMILY.
CC ---------- ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---------- ----------
DR EMBL; X62473; CAA44244.1; -.
DR PIR; S41107; S41107.
DR MGD; MGI:101815; Fmn.
DR InterPro; IPR003104; FH2.
DR InterPro; IPR001265; Formin.
DR Pfam; PF02181; FH2; 1.
DR PRINTS; PR00949; FORMIN.
DR SMART; SM00498; FH2; 1.
KW Nuclear protein; Developmental protein; Alternative splicing;
KW Phosphorylation; Coiled coil.
FT DOMAIN 418 443 COILED COIL (POTENTIAL).
FT DOMAIN 497 566 COILED COIL (POTENTIAL).
FT DOMAIN 644 748 FH1 (PRO RICH).
FT DOMAIN 759 1164 FH2.
FT DOMAIN 1043 1110 COILED COIL (POTENTIAL).
FT DOMAIN 635 638 POLY-SER.
FT DOMAIN 961 971 POLY-SER.
SQ SEQUENCE 1200 AA; 133464 MW; 4DPB1A0CB1CDB0B1T CRC64;



RESULT 11
CUT2_MOUSE
ID CUT2_MOUSE STANDARD; PRT; 1426 AA.
AC P70298;
DT 15-OCT-1996 (Rel. 34, Created)
DT 15-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2000 (Rel. 41, Last annotation update)
DE Homeobox protein Cux-2 (Cut-like 2).
GN CUTL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96394427; PubMed=8798433;
RA Quaggin S.E., Vanden Heuvel G.B., Golden K., Redmer K., Igarashi P.;
RT "Primary structure, neural specific expression, and chromosomal
RT localization of Cux-2, a second murine homeobox gene related to
RT Drosophila cut.";
RL J. Biol. Chem. 271:22624-22634(1996).
CC -!- FUNCTION: MAY BE A TRANSCRIPTION FACTOR INVOLVED IN NEURAL
CC SPECIFICATION. BINDS TO DNA IN A SEQUENCE SPECIFIC MANNER.
CC -!- SUBCELLULAR LOCATION: Nuclear.

```
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO NEURAL TISSUES. EXPRESSED
CC       EXCLUSIVELY IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.
CC   -!- SIMILARITY: CONTAINS 3 CUT DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE CUT HOMEOBOX FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U45665; AAC52762.1; -.
DR   HSSP; P10037; 1AU7.
DR   TRANSFAC; T04456; -.
DR   MEP; MEI 107321; Cut12.
DR   InterPro; IPR003350; Hmeeo_CUT.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; Homeobox; 1.
DR   Pfam; PF02376; CUT; 3.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW   Repeat; Coiled coil.
FT   DOMAIN      131    311       CUT 1.
FT   DNA_BIND    483    569       COILED COIL (POTENTIAL).
FT   DOMAIN      587    655       CUT 2.
FT   DNA_BIND    828    917       CUT 3.
FT   DNA_BIND    983   1070       HOMEOBOX.
FT   DNA_BIND   1113   1172
SQ   SEQUENCE   1426 AA;  154702 MW;  4414D5D0356B984E CRC64;

  Query Match          5.6%;  Score 129;  DB 1;  Length 1426;
  Best Local Similarity  21.7%;  Pred. No. 0.71;
  Matches  108;  Conservative   67;  Mismatches  184;  Indels  148;  Gaps   22;

Qy      27 DPKDPEALAKLGINQAHSSEPMLQPSPQKQPPPAVAPRPKAPLQLGPS---------- 73
           ||  ::  |  :  |  |  |  | | | |  |  |  |  | ||||
Db     369 DPSEDDSIKGSVGTEPPYPPQLPPPP--GPEDPLSPSPAQPL-LGPSLGPDGPRNFSLSP 425

Qy      74 ----------------------SSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPED 110
                               :: | : | | ||  |           ||  ||  :| ||
Db     426 FPSLAPGERLAGDSLLSKHMMGPAAPKGETGNLL-AFPPTFYGGAKPPSAPAASVPCPEP 484

Qy     111 PGTLSA------ -ERRCLTQPVEDQGVSTQLLAPS-GSVCFSY---------------TG 147
           |  |  |      |  :  |  |  | | ||| : |  |  |
Db     485 TGAPEAVDGASPEEEQLDTAPIAFQ VKEQLLYHNIGQRVFGHYVLGLSQGGVSEILAPP 543

Qy     148 TPW-KLFLR-KEVFYPRENF  SHPYYLRLLCEQILR-- --------DTFSESCIR-IS 191
           || || : || |   |  |     |  |  |  :  |           | |:  |: |
Db     544 KPWPKLTVVGKEPFFIKMKQFLSDEQNVLALRTIQVRQRGSITPRIRTPETGSDDAIKSIL 603

Qy     192 QNEFPFNPDLLCG ---------LEVDLDSLTTTEDSVKKPIVVAARDNWANYFSRFFPV 241
           :  :::::  ||            |  | |:| :| |  | |
Db     604 PQAKKEIFSQRGRESKNSPASVSIPNGTASSSTSEDAIKNILEQARR------------ 648

Qy     242 SGESGSDVQLLAVSHRGLRLLKVTGQPGLRPDQLKILCSYSPAEVLGVECRGGSTLELSL 301
           |  : | |  |  |  |  |  |  |  |  |  :  :  |  |||   :
Db     649 --EMQAQQQAILEMESGPPGRSVPPSPPERPSPATASQNGALTCVKQEDGGGGSGSSSTV 706

Qy     302 KSEQLVLHTAPAPAIFALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIKLLPV 361
           :  ||  |  |  | |  :|  :  :| |  |  :  | |||  |  |
Db     707 QAPLAVGSPA------AFVQRIIRKVKSEIGDAGYFDHIWASDRGLLS-----  --RPY 749

Qy     362 ATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEFGLAFWDPA 421
           |:: |    |:  ||     ||       | :|  |   |  :| | |
Db     750 ASVSP--SLSSSSSYSG-------QP-----------NGRAWPRGD----EATIAPEDEA 779

Qy     422 S----EVRKMGEGQAEA 434
           :    |  ::|| :|||
Db     780 AMGEDEAPRVGELKAEA 796


RESULT 12
APG_ARATH
ID   APG_ARATH      STANDARD;      PRT;   534 AA.
AC   P40602; Q93Z14; Q9LNT8;
DT   01-FEB-1995 (Rel. 31, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Anter-specific proline-rich protein APG precursor.
GN   APG OR AT1G20130 OR T20H2.9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94004980; PubMed-8401599;
RA   Roberts M.R., Foster G.D., Blundell R.P., Robinson J.W., Kumar A.,
RA   Draper J., Scott R.;
RT   "Gametophytic and sporophytic expression of an anther-specific
RT   Arabidopsis thaliana gene.";
RL   Plant J. 3:111-120(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed-11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   SSP consortium (Salk/Stanford/PGEC).";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
CC       TYPES IN THE ANTHER, ONLY IN MALE FERTILE PLANTS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC       MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC       MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC       MATURATION.
CC   -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC   -!- CAUTION: Ref 2 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; X03477; CAA42925.1; -.
DR   EMBL; AF022472; AAB79900.1; ALT_SEQ.
DR   EMBL; AF078847; AAL24235.1; -.
DR   PIR; S21961; S21961.
DR   InterPro; IPR001087; Lipase_GDSL.
DR   Pfam; PF00657; Lipase_GDSL; 1.
DR   PROSITE; PS01098; LIPASE_GDSL_SER; 1.
KW   Signal.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36    534       ANTER-SPECIFIC PROLINE RICH PROTEIN APG
FT   ACT_SITE    211    211       BY SIMILARITY.
FT   ACT_SITE    511    511       POTENTIAL.
FT   CONFLICT     77     77       S -> P (IN REF. 1).
FT   CONFLICT    141    141       E -> A (IN REF. 1).
FT   CONFLICT    325    325       E -> H (IN REF. 1).
SQ   SEQUENCE   534 AA;  59007 MW;  8A81BD2C8F742958 CRC64;

  Query Match             5.5%;  Score 126.5;  DB 1;  Length 534;
  Best Local Similarity   34.6%;  Pred. No. 0.31;
  Matches   37;  Conservative    9;  Mismatches   39;  Indels   33;  Gaps    6;

Qy        5 PFSPVTVTTPVQPCRPFPAPLRKIL[A]SALAPLSIPVAR-SRPMLPSPPPPAVAPPT 58
            |  | | |: |  |          |:  : :|   | :    :||   |||   |||   |:|
Db       50 RPTTQPWPNNTPTP        DPGPKPVAPPGPSGAPVAPI  JISPJPSIPI -PPQL 97

Qy       59 PAPDQLGPLLDLPRRAQPLLLPSQPLPIAHTPPP  PPAPPLPLPE 109
            |  ||          |  |:   |||  || | | |:
Db       98 KPPPAPJPJPKPJPPPKP       QPKPV   PPPACPPTPPKPQPK 135


RESULT 13
GAG_SMSAV
ID   GAG_SMSAV      STANDARD;      PRT;   512 AA.
AC   P03336;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   GAG polyprotein [Contains: Core protein P15, Inner coat protein P12;
DE   Core shell protein P30; Nucleoprotein P10].
GN   GAG.
OS   Simian sarcoma virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus
OX   NCBI_TaxID=11917;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83114004; PubMed=6298772;
RA   Devare S.G., Reddy E.P., Law J.D., Robbins K.C., Aaronson S.A.;
RT   "Nucleotide sequence of the simian sarcoma virus genome:
RT   demonstration that its acquired cellular sequences encode the
RT   transforming gene product p28sis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:731-735(1983).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licensing agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ----------------------------------------------------------------------
DR   EMBL; V01201; CAA24514.1; ALT_INIT.
DR   PIR; A33908; FOMVGS.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR000936; Gag_p30.
DR   Pfam; PF01140; Gag_MA; 1.
DR   Pfam; PF02093; Gag_p30; 1.
KW   Coat protein; Core protein; Nucleoprotein; Polyprotein.
FT   CHAIN         1    128       CORE PROTEIN P15.
FT   CHAIN       129    196       INNER COAT PROTEIN P12.
FT   CHAIN       197    455       CORE SHELL PROTEIN P30.
FT   CHAIN       456    512       NUCLEOPROTEIN P10.
SQ   SEQUENCE   512 AA;  56893 MW;  AAD8AFA4EAED7FDC CRC64;

  Query Match             5.5%;  Score 126;  DB 1;  Length 512;
  Best Local Similarity   21.0%;  Pred. No. 0.32;
  Matches  101;  Conservative   55;  Mismatches  142;  Indels  102;  Gaps   24;

Qy       12 PVQPSRPPPAPLPPIDPPLEALAELGINGAHSSPPMLSPSPGKSPPTAVAPPPKAPLQLP 71
            |  | |||       |  |  : |  |       | ||           |:|  |  |: |:
Db      119 PSAPPRPP------IYPATDDLLLL     - SEPP-     -   IYPAALPPPLAPTAVG 157

Qy       72 PPPRIPPPQPPLIPLPLPLPIAHTPPPPAPPLPLPPPPCPLCAPPPCLTQPVPPQVV  128
            |:                          |  ||    || :   :  :|    | :: |
Db      158 PA---------- - ------ ----  PGQAPDSSDPPEGPAACTPSPRA PSPADDSGPP 194

Qy      131 STQLL --       APQQQVSPLVLGIPWPLIPPEVIIPLPPSPVIP ----  LPLLP 178
            || :|          :| |    |    |  |   :::  :  :  :  :    :  '
Db      195 STVIIPIPAIGPPAEPNGLVPLQY--  -WP PSGALLYNWPLN  PPTPGNPACHTPLL 244

Qy      179 PQILPLTPP  -- PPIPIPPQNPPPPMPPLPPPPVPLPPPPPPPPPPPPRIVVAAP  228
            | ::   ||      : | ::                    | | | | :  ::||:: ||
Db      245 PSLM   PSHQPIWNPPQQI- --   ---  --     PQIIPTPPP  PPPILLPAAPPN 281

Qy      229     PNWA    - DYPPPPPVVPP      PPPPVQLLAPPPPPL                 b 266
                || |      |  :  ||::        : :  :|| | |  : |
Db      282 VLNPNGAPPLPPNLLPNPAPPLNPPLPWPYPPAAPPPLLPYPPPPVPAPPPPPPAPP 341

Qy      267 LLPVTPPPPPPPPPLLPPPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  326
            : :| |||   |          | | |    | : | : | : | : | : | : | :
Db      342 VPPVLQPPAPPPP          VPLPPPPPPAPPPYPPPPPPPPPPPPPPAAPPPAPPQPAPPP 401

Qy      327 --KARAIPALVPLPLNPLPKPDSGYVIALPSYITPNPSLLSPPPPGS                 384
              : :| | :   | :| :::  |                 : :  :
Db      402 PPPLQPLPPLQPYSLQPLVPPAPPPV---  ---          YPPPPPTPPPPLPPPPPPAPP 448

Qy      385 ----------LPLLLPVATLPPG---WQPGSAGPPSGLPPAPPLVQPAAAPPPSPPPPPQ 434
                     | ::|     | |    : |: | |:    | |  |    |    :: ||:
Db      449 PPPPPPPPPQPPPNLTPPLAAVVSPPGSGPPPPTGNLPPNPAPPTPPPPPPPPPLLPPPPQ 508

RESULT 14
PC15_HUMAN
ID   PC15_HUMAN     STANDARD;      PRT;  1955 AA.
AC   Q96QU1; Q96QT8;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Protocadherin 15 precursor.
GN   PCDH15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=21290012; PubMed=11487575;
RA   Alagramam K.N., Yuan H., Kuehn M.H., Murcia C.L., Wayne S.,
RA   Srivastava C.P.S., Kaye P.P., Kaiser P., Van Laer L., Bernier F.P.,
RA   Schwartz C., Lee C., Morton C.C., Mullins R.F., Ramesh A.,
RA   Van Camp G., Hageman G.S., Woychik R.P., Smith R.J.H.;
RT   "Mutations in the novel protocadherin PCDH15 cause Usher syndrome type
RT   1F.";
RL   Hum. Mol. Genet. 10:1709-1718(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21303027; PubMed=11398101;
RA   Ahmed Z.M., Riazuddin S., Bernstein S.L., Ahmed Z., Khan S.,
RA   Griffith A.J., Morell R.J., Friedman T.B., Riazuddin S., Wilcox E.R.;
RT   "Mutations of the p [illegible] PCDH15 cause Usher syndrome type
RT   1F.";
RL   Am. J. Hum. Genet. 69:25-34(2001).
CC   -!- FUNCTION: Calcium dependent cell adhesion protein. Essential for
CC       maintenance of normal retinal and cochlear function.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).
```

```
CC   -!- TISSUE SPECIFICITY: Expressed in brain, lung, kidney, spleen and
CC       testis. Found also in the inner and outer synaptic layers, and the
CC       nerve fiber layer in adult and fetal retinas. Found in the
CC       supporting cells, outer sulcus cells and spiral ganglion of fetal
CC       cochlea.
CC   -!- DISEASE: Defects in PCDH15 are the cause of Usher syndrome type 1F
CC       (USH1F). Usher syndrome type 1 is an autosomal recessive disorder
CC       characterized by congenital profound sensorineural hearing loss,
CC       constant vestibular dysfunction and prepubertal onset of retinitis
CC       pigmentosa leading to blindness. Usher syndrome is the most common
CC       cause of combined deafness and blindness in developed countries.
CC   -!- SIMILARITY: CONTAINS 11 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY029205; AAK31781.1; -.
DR   EMBL; AY029237; AAK31804.1; -.
DR   Genew; HGNC:14674; PCDH15.
DR   MIM; 605514; -.
DR   MIM; 602803; -.
DR   InterPro; IPR002126; Cadherin
DR   Pfam; PF00028; cadherin; 10.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 10.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal; Multigene family; Vision; Deafness; Retinitis pigmentosa;
KW   Usher syndrome.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27   1355       PROTOCADHERIN 15.
FT   DOMAIN       27   1376       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1377   1397       POTENTIAL.
FT   DOMAIN     1398   1955       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       40    147       CADHERIN 1.
FT   DOMAIN      148    265       CADHERIN 2.
FT   DOMAIN      278    395       CADHERIN 3.
FT   DOMAIN      396    509       CADHERIN 4.
FT   DOMAIN      510    616       CADHERIN 5.
FT   DOMAIN      [illegible]      CADHERIN 6.
FT   DOMAIN      719    819       CADHERIN 7.
FT   DOMAIN      820    926       CADHERIN 8.
FT   DOMAIN      927   1035       CADHERIN 9.
FT   DOMAIN     1037   1144       CADHERIN 10.
FT   DOMAIN     1145   1259       CADHERIN 11.
FT   DOMAIN     1435   1443       POLY-PRO.
FT   DOMAIN     1746   1752       POLY-PRO.
FT   DOMAIN     1815   1823       POLY-PRO.
FT   DOMAIN     1828   1831       POLY-PRO.
FT   CARBOHYD     52     52       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    201    201       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    559    559       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    662    662       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    724    724       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    768    768       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    821    821       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    951    951       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1064   1064       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1084   1084       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1175   1175       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    261    261       S -> L (IN REF. 2).
FT   CONFLICT    467    467       L -> Q (IN REF. 2).
FT   CONFLICT    929    929       Q -> R (IN REF. 2).
FT   CONFLICT   1276   1276       R -> Q (IN REF. 2).
SQ   SEQUENCE   1955 AA;  216027 MW;  5692CF30F84FF082 CRC64;

  Query Match          5.5%;  Score 125.5;  DB 1;  Length 1955;
  Best Local Similarity   22.2%;  Pred. No. 1.8;
  Matches   48;  Conservative    8;  Mismatches   54;  Indels   39;  Gaps    7;

Qy        6 PGPVPVPVQPSPPPKAPLEKIDPKDEALAKDGINSAHSSFTMLSPSPSPGPPPAVAPPPK 65
            | | | |: |  ||         |   ||              ||.||      |||:    |
Db     1746 PLPPPPPISPPSPP--------PAPAPLA---------PPPDISPFSLFCPPPS--  PPS 1785

Qy       66 APLQLGPS-------SSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAER 118
            || | |         |:      |||  |   |     |||||: | | |      ||| |
Db     1786 IPLPLPPPTFPPLSVSTSGPPTPPLLPPFPTPL- -- PPPPPSIPCPPPPSASPLSTEC 1840

Qy      119 RCLTQPVEDQGVS-TQLLAPSGSVCFSYT 146
             |:|      ||  |  | |:  :  | |
Db     1841 VCIT------GVKCTTNLMPAEKIKSSMT 1863


RESULT 15
WAS2_HUMAN
ID   WAS2_HUMAN      STANDARD;       PRT;   497 AA.
AC   Q9Y6W5; Q9UDY7; O60794;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Wiskott-Aldrich syndrome protein family member 2 (WASP family protein
DE   member 2) (Verprolin homology domain-containing protein 2).
GN   WASF2 OR WAVE2 OR SCAR2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell lymphoma;
RX   MEDLINE=99310608; PubMed=10381382;
RA   Suetsugu S., Miki H., Takenawa T.;
RT   "Identification of two human WAVE/SCAR homologues as general actin
RT   regulatory molecules which associate with the Arp2/3 complex.";
RL   Biochem. Biophys. Res. Commun. 260:296-302(1999).
RN   [2]
RP   SEQUENCE OF 2-497 FROM N.A.
RA   Machesky L.M., Insall R.H.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Bird C.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DOWNSTREAM EFFECTOR MOLECULES INVOLVED IN THE
CC       TRANSMISSION OF SIGNALS FROM TYROSINE KINASE RECEPTORS AND SMALL
CC       GTPASES TO THE ACTIN CYTOSKELETON.
CC   -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES WITH STRONGEST
CC       EXPRESSION IN PLACENTA, LUNG, AND PERIPHERAL BLOOD LEUKOCYTES, BUT
CC       NOT IN SKELETAL MUSCLE.
CC   -!- DOMAIN: BINDS ARP2/3 COMPLEX THROUGH THE C TERMINAL REGION AND
CC       ACTIN THROUGH VERPROLIN HOMOLOGY (VPH) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 WH2 DOMAIN
CC   ---------- - - --------- ----------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB026542; BAA81795.1; -.
DR   EMBL; AL022578; CAA18609.1; -.
DR   EMBL; AF134304; AAD33053.2; -.
DR   Genew; HGNC:12733; WASF2.
DR   MIM; 605875; -.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
```

```
DR   SMART; SM00246; WH2; 1.
KW   Actin binding.
FT   DOMAIN       268    271       POLY-PRO.
FT   DOMAIN       318    328       POLY-PRO.
FT   DOMAIN       333    338       POLY-PRO.
FT   DOMAIN       347    350       POLY-PRO.
FT   DOMAIN       364    371       POLY-PRO.
FT   DOMAIN       379    382       POLY-PRO.
FT   DOMAIN       391    405       POLY-PRO.
FT   DOMAIN       435    452       WH2.
FT   CONFLICT       1      1       M  > V (IN REF. 3).
FT   CONFLICT      64     64       P  > R (IN REF. 1).
FT   CONFLICT      74     74       G -> R (IN REF. 1).
FT   CONFLICT      93     93       V -> G (IN REF. 1).
FT   CONFLICT     144    144       T -> K (IN REF. 1).
FT   CONFLICT     162    162       K -> E (IN REF. 1).
FT   CONFLICT     169    169       R  > K (IN REF. 1).
FT   CONFLICT     206    206       I -> M (IN REF. 1).
FT   CONFLICT     221    222       PF -> TS (IN REF. 1).
FT   CONFLICT     250    250       S  > P (IN REF. 1).
FT   CONFLICT     311    312       SS -> GP (IN REF. 1).
FT   CONFLICT     339    339       I -> V (IN REF. 1).
FT   CONFLICT     352    352       S -> P (IN REF. 1).
FT   CONFLICT     367    367       L  > P (IN REF. 1).
FT   CONFLICT     388    388       R  > G (IN REF. 1).
FT   CONFLICT     399    399       P  > PP (IN REF. 1).
FT   CONFLICT     415    415       V  > I (IN REF. 1).
FT   CONFLICT     432    432       I -> V (IN REF. 1).
FT   CONFLICT     436    436       H -> R (IN REF. 1).
FT   CONFLICT     444    444       C -> R (IN REF. 1).
FT   CONFLICT     456    456       Q -> R (IN REF. 1).
FT   CONFLICT     461    461       W -> R (IN REF. 1).
FT   CONFLICT     492    492       G  > E (IN REF. 1).
FT   CONFLICT     497    497       N  > D (IN REF. 1).
SQ   SEQUENCE   497 AA;  54094 MW;  5F31B598FDF63A6B CRC64;

  Query Match          5.9%;  Score 125;  DB 1;  Length 497;
  Best Local Similarity   22.0%;  Pred. No. 0.35;
  Matches   54;  Conservative   30;  Mismatches   86;  Indels   76;  Gaps   11;

Qy       1 SRPQAVVVPVQPSKPPKAPLKRIDPKIEALAKLGINGAHSSPPMLSPS----PGKGPPPA 59
           :::|:  ' :  |  ||    :  | |        :|     : ||  |||    |    ||
Db     412 SKPSPAPPPAPPPPPPP--MIGIPPPP---PPIGFGSPGTPPPPSSPSFPPHPDFAAPPP 466

Qy      60 VAPPPKAPLQLGPSSSIFEFQGPIIIDLPSQHLPIAHIPPPPPAPPLPLPEDPGTLSAEPF 119
           :  : :  :     |   : :              |    ||||| ||    |  | | :  :
Db     467 LPPPAADYPILPPPPLSQ------------PTRGAPPPPPPPPPGPPPPPFTGADGQP 513

Qy     120 GLIQPVEVQGVSTQLLAPSGSVCPSYTSTPWPLPPFVPYPPENFSHPYYLPLLCEQI- 178
           :  :: :     |  |:                |   : :|    |  :|:
Db     514 AVIPPLSD-  ----TTKPKSSL-----------------PAISDAHSDLLSAICQGFQ 548

Qy     179 LPQTPSPSCIPISQNRPPKMPDILGGIFVDIDSLTTTEDSVKPRIVVAAP---------- 228
           :         ;: : : :  ;::|    |: ::      : :|| |
Db     549 LP-       -RVEPQQEQEKWDVVGN   DVATI-----LSRRIAVEYSDSEDDSSEFD 591

Qy     229 -DNWAN 233
            |:::|
Db     592 GDTWSN 597


Search completed: January 10, 2003, 02:19:37
Job time : 21.042 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

SM protein - protein search, using sw model

Run on: January 10, 2003, 01:56:04 ; Search time 28.3367 Seconds
(without alignments)
3177.594 Million cell updates/sec

Title: US-09-803-126-4
Perfect score: 2287
Sequence: 1 MYQSRPGFVPVFVQPSRDFF..... WDPASFVPKMGEGQAEARPA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2287	100.0	1560	4	Q96JP2	Q96jp2 homo sapien
2	304.5	13.3	2424	5	Q9VZ48	Q9vz48 drosophila
3	234	10.2	856	4	Q8TEE1	Q8tee1 homo sapien
4	227	9.9	1254	5	Q8T0A9	Q8t0a9 drosophila
5	227	9.9	2167	5	Q9V3Z6	Q9v3z6 drosophila
6	215.5	9.4	2179	13	Q9DGG9	Q9dgg9 brachydanio
7	215.5	9.4	2179	13	Q9DGG8	Q9dgg8 brachydanio
8	208.5	9.1	2113	11	Q99MZ6	Q99mz6 mus musculu
9	204.5	8.9	2810	5	Q20456	Q20456 caenorhabdi
10	200	8.7	921	11	Q8R0U3	Q8r0u3 mus musculu
11	182	8.0	2098	5	P91443	P91443 caenorhabdi
12	148.5	6.5	555	12	Q9YMX1	Q9ymx1 lymantria d
13	142.5	6.2	559	11	Q91W51	Q91w51 mus musculu
14	141.5	6.2	559	11	Q9ERQ9	Q9erq9 mus musculu
15	141.5	6.2	559	11	Q8R5H6	Q8r5h6 mus musculu
16	140	6.1	493	3	Q9UVD1	Q9uvd1 pneumocysti
17	138	6.0	525	6	Q95L53	Q95153 mustela vis
18	137.5	6.0	2121	5	Q9NH54	Q9nh54 drosophila
19	137	6.0	938	3	Q96VI4	Q96vi4 pneumocysti
20	136.5	6.0	409	10	Q9SBM1	Q9sbm1 solanum tub
21	136.5	6.0	2129	5	Q9VL23	Q9vl23 drosophila
22	136	5.9	263	17	Q8ZYD1	Q8zyd1 pyrobaculum
23	136	5.9	559	4	Q8WV09	Q8wv09 homo sapien
24	135	5.9	1420	3	O13736	O13736 schizosacch
25	134	5.9	1217	4	Q9ULL5	Q9ul15 homo sapien
26	133	5.8	525	10	Q8W158	Q8wi58 brassica ol
27	133	5.8	699	10	O81765	O81765 arabidopsis
28	132	5.8	893	3	Q8WZS5	Q8wzs5 neurospora
29	132	5.8	1059	10	Q94E61	Q94e61 oryza sativ
30	131.5	5.7	282	4	Q9BSA1	Q9bsa1 homo sapien
31	131.5	5.7	1134	4	Q96JH1	Q96jh1 homo sapien
32	130.5	5.7	566	6	Q95JD1	Q95jd1 sus scrofa
33	130.5	5.7	1154	4	Q9P2P0	Q9p2p0 homo sapien
34	130.5	5.7	1201	10	Q9C6S1	Q9c6s1 arabidopsis
35	130.5	5.7	1431	4	Q8WXD9	Q8wxd9 homo sapien
36	130	5.7	511	6	Q95JD0	Q95jd0 sus scrofa
37	130	5.7	960	10	Q8S0F0	Q8s0f0 oryza sativ
38	129.5	5.7	472	3	O59907	O59907 neurospora
39	129	5.6	446	6	Q97727	Q97727 sus scrofa
40	129	5.6	643	6	Q97726	Q97726 sus scrofa
41	129	5.6	3105	12	Q8UZI9	Q8uzi9 cercopithic
42	128.5	5.6	802	11	P70433	P70433 mus musculu
43	128	5.6	451	11	Q62775	Q62775 rattus norv
44	128	5.6	485	11	Q9Z0G8	Q9z0g8 rattus norv
45	128	5.6	727	10	Q9XIL9	Q9xil9 arabidopsis

ALIGNMENTS

RESULT 1
Q96JP2
ID Q96JP2 PRELIMINARY; PRT; 1560 AA.
AC Q96JP2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1783 protein (Fragment).
GN KIAA1783.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21245133; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large Proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058686; BAB47412.1; -.
DR InterPro; IPR000048; IQ_region
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR001452; SH3
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MyTH4; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50096; IQ; 1
DR PROSITE; PS50002; SH3; 1
FT NON_TER 1 1
SQ SEQUENCE 1560 AA; 170146 MW; 6A61734C61E4C35A CRC64;

Query Match 100.0%; Score 2287; DB 4; Length 1560;
Best Local Similarity 100.0%; Pred. No. 1.1e-172;

```
  Matches  437;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MVQVRLQDVTVDVSVSLQEFRAPFPSEFPAALAPLIDTAHLSFMLSESGPGPEAV 56
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1124 MVQVRLQDVTVDVSVSLQEFRAPFPSEFPAALAPLIDTAHLSFMLSESGPGPEAV 1179

Qy       57 APPDVAPLQDGPSSLIPEFQGELLELFQKRIEAMCPEFPAPLLLELDTGILSAEPP 112
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1180 APPDVAPLQDGPSSLIPEFQGELLELFQKRIEAMCPEFPAPLLLELDTGILSAEPP 1235

Qy      113 LTQPVEPWCKSTTLIAFVVLFPSYTSTFWFLFEFFVFYFFNFCHFYYLEFLLCFQILF 169
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1236 LTQPVEPWCKSTTLIAFVVLFPSYTSTFWFLFEFFVFYFFNFCHFYYLEFLLCFQILF 1292

Qy      170 VTPVPNVIPRTQNPFEHVLLVSRLVTIFCLTTTECVFFFIVVAARDNWAVYFSFFF 226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1293 VTPVPNVIPRTQNPFEHVLLVSRLVTIFCLTTTECVFFFIVVAARDNWAVYFSFFF 1349

Qy      227 YSFSGSVRIGQNPFEHVLLVSRLVTIFCLTTTECVFFFIVVAARDNWAVYFSFFF 283
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1350 YSFSGSVRIGQNPFEHVLLVSRLVTIFCLTTTECVFFFIVVAARDNWAVYFSFFF 1406

Qy      284 YSFSQLVLHTAPAPAIEALVELFLNELKFFCQYVIALPCYITMCFLLGFHPCDLIKLLP 343
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1407 YSFSQLVLHTAPAPAIEALVELFLNELKFFCQYVIALPCYITMCFLLGFHPCDLIKLLP 1466

Qy      344 VALSFSWTDFFYVVTKAWLALGVLAAWLLSIGEPGLARWSP 403
            |||||||||||||||||||||||||||||||||||||||||||
Db     1467 VALSFSWTDFFYVVTKAWLALGVLAAWLLSIGEPGLARWSP 1526

Qy      404 ASEVEFMGEGQAEAPPA 437
            |||||||||||||||||
Db     1527 ASEVEFMGEGQAEAPPA 1560


RESULT 2
Q9VZ48
ID   Q9VZ48      PRELIMINARY;      PRT;  2424 AA.
AC   Q9VZ48;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG2174 protein.
GN   CG2174.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Holometabola; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20076886; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., Mi-ll- E.M.E., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu Y., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003494; AAF47980.1; -.
DR   HSSP; P08799; 1MND.
DR   FlyBase; FBgn0030252; CG2174.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000255; myosin_head; 1.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 1.
SQ   SEQUENCE   2424 AA;  267616 MW;  8AD67AA3FF9AA0C1 CRC64;

  Query Match             13.8%;  Score 304.5;  DB 5;  Length 24.4;
  Best Local Similarity   33.7%;  Pred. No. 5.4e-14;
  Matches   89;  Conservative   98;  Mismatches   90;  Indels   41;  Gaps   1;

Qy       89 QKLPIAH TPPPPPAPPLPIPEPPPSTLSAEPPVCIPQPVPPQPVSVSPLIAPPQS   VTP 144
            ||:   | | |||   | |:|   |                 ||| :  :| |     |
Tl     2127 PEIPMHMTTPPIVIPAPSTIPTPT                 PPPTPPPMPPPGGPAPL 2176

Qy      145 SYTGTPWKLFLRREVEYPPRENPSHPYYIKLLEQLIRPIPSSR  PPINLNERFPMKLLL 202
            :|   |||| :|||| | |     | | ||  :| | |  : :::  :|     :|
Db     2177 TYNKVPWFIRVPREVEVPQHEPIGPPVALIELLPALVDIPVPLP LPIIPLERSPAINMI 2236

Qy      203 GSIFVLITSLTTTFIG -            VPPPIVVAAGTWAALPPPPPVGMSGGGGL 247
            | | :|:|:  |                  ||: :|  ||| :  |:: | |  |
Tl     2237 HSHEVPVTIPPNPMSE SAISSSPS PAPVPPPNAVMAP  MPIPPAPIPAVGMAPPL 2293

Qy      248 PVILLAVDHPLPLLPVTLSPLPFLLPKLLPVPGIAIT LPPLPPPPPPPPPPPPPPPP 307
            || :: ||| || | :         | | :: : || |:   |     : |||||| :
Db     2338 PVSIMSVPHSGLYIAP      PPAPTLIVVQAISPSEIQSAVTIIPTPAAILINLPNSPH 2391

Qy      308 LVIHTAPAPAIPALVPLPLNPIKK 329
            | || ||| ||:::|  |: | :|
Db     2392 IAIHAAPAPAAAIQSMVTIPVLPPPK 2415


RESULT 3
Q8TEE1
ID   Q8TEE1      PRELIMINARY;      PRT;  856 AA.
AC   Q8TEE1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FLJ00256 protein (Fragment).
GN   FLJ00256.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
P    SEQUENCE FROM N.A.
C    TISSUE=SPLEEN;
A    Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Chara O.;
T    "The nucleotide sequence of a long cDNA clone isolated from human
T    spleen.";
L    Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
R    EMBL; AK074183; BAB85009.1; -.
T    NON_TER       1       1
Q    SEQUENCE   856 AA;  96020 MW;  A2B5C5AD6B3C1B53 CRC64;

  Query Match              10.2%;  Score 234;  DB 4;  Length 856;
  Best Local Similarity    24.2%;  Pred. No. 5.3e-10;
  Matches  116;  Conservative   63;  Mismatches  157;  Indels  144;  Gaps   21;

Qy      31 EALAKLGINGAHSSPPMLSPPGPGKGPPPAVAPPPKAPLQL------------GPSSS-- 75
           | | |:  :|   :|:|                  ||  :  |:|:                |.|
Db      67 EPLPPTYANGVKAEPPTWLEL---  --QAVPSKKHIPIQVILATGESLTVPVDSASTSPE 120

Qy      76 -  -IKEKGGPLLDLFGQKLPIA------------HTPPPPPAPPLPLPEDPGTLSAER 118
           |  |||| | |  |  | |:|                                  :|
Db     121 MQMHIAHKQG-LSDHLGFSLQVAVYDKPWSLGSGRDH---  ------------MMDAIA 162

Qy     119 PGLTQPVPEQGVSTQLLAPGGSVCFSYTGTPWKLFLRKEVFYP-RENFSHPYYLPLLCPQ 177
           ||  |  |:|| | :                  :|| | | |  |  |  |  |:  |  |
Db     163 PG-PQMAQPPGESQR------------QSPWRIYFRKEFFTPWHDSREDPVSTELIYPQ 208

Qy     178 ILRDTFS--  -----ESCIRI----------SQNERRKMKDLIGG--------------- 204
           :|| |:|     |     |:|||          |  |  | | :  :|
Db     209 VLRGVWSGFYSFFKEEELVELLAPHCYVQLGASAESKAVQELLPSCIPHELYFFPPPPFW 268

Qy     205 ------  LEVDLDSLTTTEDSVKKPIVVAAPDNWANYFSPPFPVSGESGSDVQ----LL 252
                   :  |  |  |:|:|:|  |||  |  |||  | |  ||  :   :|
Db     269 ASLVTAACAKAPYTQKQVTPLAVREQVVDAARLQWPLLFSRLFEVITLSGHRLPPTQLLL 328

Qy     253 AVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGV----ECRGGSTLELSLKSEQLVL 308
           ||:  :||  |           | |:|  || ||:|:    | :||  | ||   |:
Db     329 AVTWKGLGFL  ------DQQEKMLLELSFPEVMGLATNREAQGGQRLLLSTMHEEYEF 379

Qy     309 HTARAPAIEALVELFLNELKKDSGYVIALRS-YITDNCSLLSFHRGDLIKLLPVATL--E 365
           :  : ||  || |||  ||: | : :||:   ||: :|:|:| :|||: |      |
Db     380 VSPSSVAIAELVALFLFGLKERSIFAMALQDRKATDDTTLLAFKKGDLLVLTKKQGLLAS 439

Qy     366 PGWQFG  PAGGPSGLFP-------ADIVQPAA--APDFSFSKEQRS-GWHKGQLSNGEP 413
           |  | |      |  ||  |         : :|:|    : | :|:|      :|| .  |
Db     440 ENWTLGQNERTGKTGLVPMACLYTIPTVTKPSAQLLSLLAMSPEKRKLAAQEGQFTEPPP 499
```

```
RESULT 4
Q8T0A9
ID   Q8T0A9      PRELIMINARY;      PRT;  1254 AA.
AC   Q8T0A9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   LD14917p.
GN   CK.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunez J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY069438, AAL39583.1; -.
SQ   SEQUENCE   1254 AA;  144367 MW;  806DEC6DBCAD23F9 CRC64;
```

```
  Query Match              9.9%;  Score 227;  DB 5;  Length 1254;
  Best Local Similarity   23.8%;  Pred. No. 3.2e-09;
  Matches  111;  Conservative   53;  Mismatches  154;  Indels  148;  Gaps   19;

Qy      20 EAFLPYTDPY-----DEALAKLGINGAHSSPP MLSPSPGKGPPPAVAP  - -  PPKAP 67
           .|:|:|  |       :| | :   ||  : || |     |  | :|  |       |
Db     292 PAPIREGPPGYAPYCEERLKRTFNNGTENQFTSWLELQATYSKKPIMLPITFMDSNTKTL 351

Qy      68 L------------QLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLS 115
           |            ||   |:|       | ||  | ||
Db     352 LADSATTARELCNQLSDKISLK-------DQFGFSLYIA---------------------- 383

Qy     116 AERRCLTQPVEDQGVSTQLLAPSGSVCFSYT--------GTPWKLFLRKEVFYPRENFSH 167
           |  |  |  |   :  : | | |              ||:|| |||:| |   :|
Db     384 -----LFDKVSSLGSGGDHVMDAISQCELYAPEQPAIPENAFWRLFPPKEIPAPWHEPTH 438

Qy     168 YYY LPLLCPQILRDTFSESCIRISQNERRKMKDL------------------------ 201
               |  |:|| |  :  :   | :||
Db     439 DQVATNLIYQLVVPG-  ---VKFGKYFCDKEEDLAMIAAQQTFIEYSTDMSMFRLFTLL 492

Qy     202 -------LGGLEVDLDSLTT-----------TED- ---SVKKPIVVAARDNWANYFSR 237
                  | |..  .:.                |      :|: ||  |:  |  |||
Db     493 PNFIPDFCLSGVDKAIERWAALVLQAYKKSYYVYDFIAPLKIKEDIVCYAKYYWPLLFSR 552

Qy     238 PFPVSGESGSDVQ----LLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECRG 293
           |:    || ::|                :|:  |:                || :|:|  || |:  |
Db     553 FYEAYRNSGPNLPKNDVIIAVNWTGVYVV---  ---DLQEQVLLELSFPEITAVGLCP 608

Qy     294 GS-----TLELS-LKSEQLVLHTAPARAIEALVELFLNELKKDSGYVIALRST  TTENS 345
           .  |     | |:  :|:  .  |  |  | | |  ||  ||: ||| | |||||: |  :|
Db     609 TNKVPTQTPSLSTVRGEEFTFQSPNAEDIEDLVVYFLDSLYFPSPYVIAIQPYPAPSHGT 668

Qy     346 SLLSFHRGDLIKL----LPVATLEPGWQFG--SAGGRSGLFPADIV 385
           | |||  :|||  |        :|  ||  |         | |||: |
Db     669 SFLSFFKGDLIILEDESCGESVLNNGWCIGPMTRSQEEGDFTAETV 709
```

RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.C., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.",
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN BERKELEY;
RX MEDLINE=99403001; PubMed=10471707;
RA Adams M.D., Myers E.W., Smith H.O., Lewis G.F., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Venter J.C., Rubin G.M.;
RT "Annotation of the sequence of 2.9 Mb region of the genome of
RT Drosophila melanogaster: the Adh region.",
RL Genetics 153:179-219(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN BERKELEY;
RA Svirskas R.R., Agbayani A., Arcuina T.T., Baxter E., Blazej R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan T.F., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.P., Farra F., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomotan M.A., Mazda F., Moshrefi A.F., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Pun L., Sequeira A.,
RA Sethi H., Shir E., Svirskas P.R., Wan K.H., Weinburg T., Zhang P.,
RA Zlotan L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF003646; AAP53435.1; -.
DR EMBL; AF003412; AAF44915.1; -.
DR HSSP; P08799; 1MND.
DR FlyBase; FBgn0020317; ck.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR000567; SBP_bac_1.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 4.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MyTH4; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000365; myosin_head; 1.
DR SMART; SM00295; B41; 2.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSc; 1.
DR SMART; SM00139; MyTH4; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50057; BAND_41_3; 2.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS51037; SPP_PA_TPR_1; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 2179 AA; 249307 MW; 3CE7E24AEE1B9A42 CRC64;

Query Match 9.9%; Score 227; DB 5; Length 2179;
Best Local Similarity 28.4%; Pred. No. 6.6e-09;
Matches 111; Conservative 53; Mismatches 174; Indels 148; Gaps 17;

Qy 20 KAPLPKIDPK---- DFALAKLSINGAHRSDP MLPPPSAPKPPPAVAP 69
Db 1105 KAPIPE-DE-TAPT-EPPLIPEIPDDIPEN-I- WIEI-AIP ---L.MIIIPMI-NPPPT 1154

Qy 68 L--- -------QLGPSNSIKPPQPPIIPPPQPPIAP-PPPPAPPPPIPPPPPPP 116
Db 1155 LAI-ATTAPPIDQIPPPIPP PLPPIPLYIA - 1200

Qy 116 AERRQLTQPVEDQGVSTQLLAPSGSVTPSYT -- QTIWPLPLKREVPYTRENPGH 167
Db 1201 IPIPVP-P-P-P-PPPVMIAI-P-PIAP-P-AP-P-NAPWPIPPPRPIPAPWHEPTH 1254

Qy 168 PYY LPLLCEQILRITPCLCIAIS-NLRRFMHDL 201
Db 1255 PIVATNLIYQQVYFS VKPGEYPCTPEETLAMIAAQCVTIHYGTTPSMERLPTLL 1314

Qy 202 LGCLEVDLDDLTT TPD SVPKPPIVVAAPLNWANYPDP 237
Db 1315 PNPIPPPLPSGVPPAIPPWAALVIQAYPPYYVPPPIAPPPIPPPPPGYAPYPPWPPIPPW 1374

Qy 238 PPIVAP SPLVL IIAVSHPNRIIPVTQQPIEPTLPTIVTPAPVTPETRT 287
Db 1375 PYPAYPMQPPMPPMYIIAVPWTPYYYY PPPPVVIPPPPPPPAVPNPPK 1424

Qy 288 GL -- TLELL LPLEQLVLHTAPAPAIPAIVPIP.NPIPPL-IPIIAIIPY 334
Db 1425 TNPVPT-PPSLSTVPPEEFTPQPPAPPIPPLVYYPPPPPPPPPPPPPPPPPPPPPK 1474

Qy 346 SLLSFHRGDLIKL- -LPVAILEPQWQPG SAGGRSGLFPASIV 385
Db 1477 SFLSFPKCDLIILEDPCCGPCVLNNGWCIGPCPPSQPPPHPPAETV 1522

RESULT 6
Q9DGG9
ID Q9DGG9 PRELIMINARY; PRT; 2179 AA.
AC Q9DGG9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 21-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Myosin VIIA.
GN MYO7A.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GDS;
RX MEDLINE=20414642; PubMed=10958658;
RA Ernest S., Rauch G.J., Haffter P., Geisler R., Petit C., Nicolson T.;
RT "Mariner is defective in myosin VIIA: a zebrafish model for human
RT hereditary deafness.";
RL Hum. Mol. Genet. 9:2189-2196(2000).
DR EMBL; AJ404001; CAC05418.1; -.
DR HSSP; P08799; 1MND.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00612; IQ; 5.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MyTH4; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000365; myosin_head; 1.

```
R   SMART; SM00295; B41; 2.
R   SMART; SM00015; IQ; 4.
R   SMART; SM00242; MYSc; 1.
R   SMART; SM00139; MyTH4; 2.
R   SMART; SM00326; SH3; 1.
R   PROSITE; PS50057; BAND_4.1_3; 2.
R   PROSITE; PS50096; IQ; 2.
R   PROSITE; PS50002; SH3; 1.
Q   SEQUENCE   2179 AA;  251632 MW;  F53AE9512ADEBBFB CRC64;

 Query Match             9.4%;  Score 215.5;  DB 13;  Length 2179;
 Best Local Similarity   24.4%;  Pred. No. 5.4e-08;
 Matches  108;  Conservative   52;  Mismatches  154;  Indels  129;  Gaps   20;

Qy      30 DFALAKLGINGAHSSPPM---LSPSPGKGP --PPAVAPRPKAPLQLGPSSSIKE---- 78
           :| |  . :||  : ||    |  .  | |  |          |     ... ||
Db    1228 EERLRRTFVNGTRTQPPSWLELQATKSKPIMLPVTFMDGTTKTLLTDSATTAKELCNAL 1287

Qy      79 KQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAP 137
           :  | | ||  | ||                          |  |  |     :
Db    1288 SDKISLQDFFGFSLYIA------------------------LFDKVSSLGSGNDHVMD 1321

Qy     138 SGSVCFSYT--------GTPWKLFLPKEVFYP-RENFSHPYYLPLLCFQILRDT-FSE-S 186
           : | | |           ||:|| | | :  :          ||: :|| :|  | |  :
Db    1322 AVSQCEQYAFEQGAQFFNAPWPLFPPKEIFTPWHDPAEDAVATNLIYQQIVRGVKFGEYP 1381

Qy     187 CIRISGNERPKMKDLLG-------GLEV----------------DLDSLTTTE------- 216
           |    :|  : :|          | |:                 : | | | |
Db    1382 C------DREDLAELASQQYYVDYGSEILVERLLSLIPSYIPDREISSAKTVERWAQFIM 1435

Qy     217 ----------  - --DSVKKRIVVAARDNWANYFSRF---FPVSGES--GSDVQLLAVSH 256
                    ||  :|  | |  | |||| |  || |  .|| .:|||:
Db    1436 AAHKKGVYTQKRVDPQKVKEEVVDFARYKWPLLFSRFYEAFKFSGPSLPKNDV-IVAVNW 1494

Qy     257 RGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECRGGSTLE------LSLKSFQLVLHT 310
           |:  :          :| ..|  || |  |  |  |  |  .:|  .  :
Db    1495 TGVYFV----  ----DEQEQVLLELSFPEITAVSSSKGGKLQAQSFTLATIKADEFIFTS 1545

Qy     311 ARAPAIEALVELPLNELKKDSGYVIALR---SYITDNCSLLSFHRGDLIKL---LPVATL 364
           |  |  ||  ||  |:|  | .:|||:   |   |: : ||| :|||| |          :
Db    1546 NNAEDIRDLVVTFLEGLRKPSKFVVALQDNPSPAADDSTFLSFLKGDLIVLDQDTGEQVM 1605

Qy     365 EPGWQFGS--AGGRSGLFPADIV 385
           || |:      : | |||| |
Db    1606 TSGWAHGTNDRTKQRGDFPADCV 1628


RESULT 7
Q9DGG8
ID   Q9DGG8       PRELIMINARY;      PRT;  2179 AA.
AC   Q9DGG8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Myosin VIIA.
GN   MYO7A.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TUEBINGEN;
RX   MEDLINE=20414642; PubMed=10958658;
RA   Ernest S., Rauch G.J., Haffter P., Geisler R., Petit C., Nicolson T.;
RT   "Mariner is defective in myosin VIIA: a zebrafish model for human
RT   hereditary deafness.";
RL   Hum. Mol. Genet. 9:2189-2196(2000).
DR   EMBL; AJ404002; CAC05419.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50057; BAND_4.1_3; 2.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50002; SH3; 1.
SQ   SEQUENCE   2179 AA;  251655 MW;  FF440D1EFD588F76 CRC64;

 Query Match             9.4%;  Score 215.5;  DB 13;  Length 2179;
 Best Local Similarity   24.4%;  Pred. No. 5.4e-08;
 Matches  108;  Conservative   52;  Mismatches  154;  Indels  129;  Gaps   20;

Qy      30 DEALAKLGINGAHSSPPM---LSPSPGKGP---PPAVAPRPKAPLQLGPSSSIKE---- 78
           .| |  . :||  : ||    |  .  | |  |          |     ... ||
Db    1228 EERLRRTFVNGTRTQPPSWLELQATKSKPIMLPVTFMDGTTKTLLTDSATTAKELCNAL 1287

Qy      79 -KQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAP 137
            :  | | ||  | ||                          |  |  |     :
Db    1288 SDKISLQDFFGFSLYIA------------------------LFDKVSSLGSGNDHVMD 1321

Qy     138 SGSVCFSYT-------GTPWKLFLPKEVFYP-RENFSHPYYLRLLCEQILRDT-FSE-S 186
           : | | |           ||:|| | | :  :          ||: :|| :|  | |  :
Db    1322 AVSQCEQYAFEQGAQERNAPWPLFPPKEIFTPWHDPAEDAVATNLIYQQIVRGVKFGEYP 1381

Qy     187 CIRISGNERPPMKDLLG-------GLEV----------------DLDSLTTTE------- 216
           |    :|  : :|          | |:                 : | | | |
Db    1382 C------DREDLAELASQQYYVLYGSEILVERLLSLIPSYIPDREISSAKTVERWAQFIM 1435

Qy     217 ----------DSVKKRIVVAARDNWANYFSRF---FPVSGES--GSDVQLLAVSH 256
                    ||  :|  | |  | |||| |  || |  .|| .:|||:
Db    1436 AAHKKGVYTQKRVDPQKVKEEVVDFARYKWPLLFSRFYEAFKFSGPSLPKNDV-IVAVNW 1494

Qy     257 PGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECRGGSTLE---   LSLKSFQLVLHT 310
           |:  .          :| ..|  || |  |  |  |  |  .:|  .  :
Db    1495 TGVYFV---------DEQEQVLLELSFPEITAVSSSKGGKLQAQSFTLATIKADEFIFTS 1545

Qy     311 ARARAIEALVELPLNELKKDSGYVIALR---SYITDNCSLLSFHRGDLIKL---LPVATL 364
           |  |  ||  ||  |:|  | .:|||:   |   |: : ||| :|||| |          :
Db    1546 NNAEDIRDLVVTFLEGLRKRSKFVVALQDNPSPAADDSTFLSFLKGDLIVLDQDTGEQVM 1605

Qy     365 EPGWQFGS--AGGRSGLFPADIV 385
           || |:      : | |||| |
Db    1606 TSGWAHGTNDRTKQRGDFPADCV 1628


RESULT 8
Q99MZ6
ID   Q99MZ6       PRELIMINARY;      PRT;  2113 AA.
AC   Q99MZ6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Myosin-VIIb.
GN   MYO7B
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21295050; PubMed=11401444;
```

```
RA   Chen Z.-Y., Hasson T., Zhang D.-S., Schwender B.J., Derfler B.H.,
RA   Mooseker M.S., Corey D.P.;
RT   "Myosin VIIb, a novel unconventional myosin, is a constituent of
RT   microvilli in transporting epithelia.";
RL   Genomics 72:285-296(2001).
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AF242411; AAK28339.1; -.
DR   HSSP; P08799; 1MND.
DR   MGD; MGI:107709; Myo7b.
DR   InterPro; IPR002106; AAtRNA ligaseII.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000871; Beta_lactamase_A
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR000508; SigPTase
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; [illegible]; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; [illegible]; BETA_LACTAMASE_A; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS00761; SPASE_I_3; UNKNOWN_1.
KW   SH3 domain.
SQ   SEQUENCE   2113 AA;  240785 MW;  [illegible] CRC64;

  Query Match             9.1%;  Score 208.5;  DB 11;  Length 2113;
  Best Local Similarity   23.7%;  Pred. No. 1.9e-07;
  Matches  [illegible];  Conservative  [illegible];  Mismatches  149;  Indels  149;  Gaps  20;

Qy         29 [illegible] 70
Db       1139 [illegible] 1198
Qy         71 [illegible] 98
Db       1199 [illegible] 1254
Qy         99 [illegible] 158
Db       1255 [illegible] 1286
Qy        159 [illegible] 184
Db       1287 [illegible] 1346
Qy        185 [illegible] 235
Db       1347 [illegible] 1405
Qy        236 [illegible] 289
Db       1406 [illegible] 1455
Qy        290 [illegible] 345
Db       1456 [illegible] 1515
Qy        346 [illegible] 381
Db       1516 [illegible] 1555
```

```
RESULT 9
Q20456
ID   Q20456      PRELIMINARY;      PRT;  2810 AA.
AC   Q20456;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F46C3.3 protein.
GN   F46C3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cottage A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z66563; CAA91469.1; -.
DR   HSSP; P10587; 1BR2.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 3.
DR   Pfam; PF00784; MyTH4; 2.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   PROSITE; PS50096; IQ; 1.
SQ   SEQUENCE   2810 AA;  [illegible] MW;  275EF5064B758FB6 CRC64;

  Query Match             8.9%;  Score 204.5;  DB 5;  Length 2810;
  Best Local Similarity   22.3%;  Pred. No. 5.7e-07;
  Matches  [illegible];  Conservative  62;  Mismatches  176;  Indels  [illegible];  Gaps  [illegible];

Qy         39 NGAHSSPPMLSPSPGPGPPPAVAPPPPAPLQ[illegible] 76
Db       [illegible] [illegible]
Qy         77 ----------KEKQGHLLDLFG[illegible] 116
Db       [illegible] [illegible]
Qy        117 [illegible]
Db       [illegible] [illegible]
Qy        [illegible] [illegible]
Db       [illegible] [illegible]
Qy        232 [illegible] 288
Db       [illegible] [illegible]
Qy        289 [illegible] 341
Db       [illegible] [illegible]
Qy        342 [illegible] 394
Db       [illegible] [illegible]
Qy        395 FSKEQR 400
```

```
                ::::|
Db    2267 EDRDER 2272


RESULT 10
Q8R0U3
ID   Q8R0U3      PRELIMINARY;      PRT;   921 AA.
AC   Q8R0U3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similar to myosin VIIb (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata, Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC026416; AAH26416.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   921 AA;  104379 MW;  04FE038A45B3193E CRC64;

  Query Match             8.7%;  Score 200;  DB 11;  Length 921;
  Best Local Similarity  26.9%;  Pred. No. 2.3e-07;
  Matches   78;  Conservative   42;  Mismatches  102;  Indels   68;  Gaps   11;

Qy      149 PWKLFLPKEVFYP-RENFSHPYYLRILCEQILRDTFS ---------------------- 184
            ||:::: ||| | |  ::  |    |  :|:|  |
Db       85 PWRIYFRKEFFTPWHDSQEDPVSTELIYHQVLRGVWSGEYNFEKEEELVELLARHCYVQL 144

Qy      185 -------------ESCIRISQNEPPFMKDLLGGL------EVDLDSLTTTEDSVKKRIVV 225
                        ||: |  | | |  .  |  |      :        | :|::: |
Db      145 GATVKSNAVQELLPSCVP SKLYPTKSPEKWASLVTAAHAKAQYTQSKATHLAVFPQTVE 203

Qy      226 AAPENWANYFSRFFPVSGESGSDVQ----LLAVSHRGLRLLKVTQGPGLRPDQL-KILCS 280
            |||  |   ||| | |:  ||  .      :|| .:  |  |              || : |
Db      204 AARLLWPLLFSRLFEVTTLSGPRLPKTQLVLAINWKGMYFL----------DQKEPTLLG 253

Qy      281 YSFAEVMVV-  -PTFQSSTLELGLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIA 336
            |||||:|:     :  ||  | |  .      :||  | ||  | |  |  |  |  | |
Db      254 LSFAEVMGLVANRDAFGGKKLLIATLQEFYRFVSFSSVAIAEMVALFLGGLFERSVFAMA 313

Qy      337 LRS YITDNCSLLSFHRGDLIKLLPVATL--EPGWQFG--SAGGRSGLFP 381
            |:   ||: :|| | :|||: |    |      |  |    |::|| |
Db      314 LQDRRATDDITLLPFKKGDLLILTFKQGLLASENWALGQNDRTGKTGLVP 363


RESULT 11
P91443
ID   P91443      PRELIMINARY;      PRT;  2098 AA.
AC   P91443;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   T10H10.1 protein.
GN   T10H10.1.
OS   Caenorhabditis elegans
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7907988;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Nelson J., Langston Y.;
RT   "The sequence of C. elegans cosmid T10H10.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80848; AAB37988.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000299, Band_4_1.
DR   InterPro; IPR000048, IQ_region
DR   InterPro; IPR001609; myosin_head
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF00612; IQ; 4.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00314; RA; 1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS50096; IQ; 1.
DR   PROSITE; PS50096; IQ; 1.
SQ   SEQUENCE   2098 AA;  239783 MW;  A97ZE51FC54B73BB CRC64;

  Query Match             8.0%;  Score 182;  DB 5,  Length 2098;
  Best Local Similarity  21.1%;  Pred. No. 2.4e-05;
  Matches   89;  Conservative   60;  Mismatches  152;  Indels  122;  Gaps   16;

Qy       23 FLPKIDPK--------DEALAKLGINGA-HSSPPMLSPSPGKGPPPAV---------- ---  60
            |:|.  |         |  |  |  |  |  |  |  |  |  |  |          |
Db     1127 FIPEPGPAGTGYSKYIEDRLPFTQVNGTPHQPPSTVELQANRSLPIVVLAVTPMSTPP  1186

Qy       61 --APRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAEP  118
              |     :|    :  | | | |  | | |  |  |  |  |  |  |  |  |  |
Db     1187 LCADSATTAAEL--CKQLAEKVG-LTNSPGFSLYIA----------------------  1219

Qy      119 RCLTQPVEDQGVSTQLLAPSGSVCFSYT--------GTPWKLFLRKEVFYP-RENFSHPY  169
            |  |  |  |  |  :  | | |  |                ||:|| |||:| |  :     |
Db     1220 --LFDKVSSLGSGTDHVMDAISQCEQYAKEQGRQERNAFWRLFPPKEIFSPWHDPRDDPV 1277

Qy      170 YLRLLCEQILR------------------------DTFSESCIRISQNEPPFMKDLLGGL  203
            |:  | |.|                        |  :  ::  |  :  | |  :
Db     1278 STNLIYQQVIRGIKYGEYRCDKDEELAAICAQQYYIDEJTMEQIFLENNLPSYLPDPEMS 1337

Qy      204 GLEVDLDSLTTT--------------EDSVKKPIVVAARDNWANYFSRFFPVSGESGSD  248
            | |:  |  | |                :  ||:  :|  |:  |  |  ||||:     :|
Db     1338 GKEMAIEKWTQTIMHQYRKKFTGRLPSQIEVKENVVCVAKTKWPLLFSRFYEALKFAGPP 1397

Qy      249 VQ----LLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVEQGG  STLLE  299
            :     ..||:  |                |  ..  .|| |.  |  |     | ||  :
Db     1398 LPKNEVIIAVNWTGVYVV---------DDREHVMLEFSFPEISTAYYGKFFSTTDTCYV 1448

Qy      300 -SLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITD-NCSLLSFHRGDLIK  357
             |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     1449 RTVVGDEYTFQSPNADDITNLIVMFLESLKKSSFYLVAIPWQPTFPPNPLPPPFGDLLI 1508

Qy      358 LL 359
            ||
```

```
Db    1569 LV 1610


RESULT 12
Q9YMX1
ID   Q9YMX1      PRELIMINARY;      PRT;   555 AA.
AC   Q9YMX1;
DT   31 MAY 1999 (TrEMBLrel. 10, Created)
DT   31 MAY 1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Essential structural protein pp78-81.
OS   Lymantria dispar multicapsid nucleopolyhedrosis virus (LdMNPV).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=10449;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99174785; PubMed=9887315;
RA   Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA   Slavicek J.M., Rohrmann G.F.;
RT   "Sequence and analysis of the genome of a baculovirus pathogenic for
RT   Lymantria dispar.";
RL   Virology 253:17-34(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kuzio J.;
RL   Submitted (MAR 1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF081810; AAC19811.1;  -.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00246; WH2; 3.
DR   PROSITE; PS00047; MYB_1; UNKNOWN_1.
SQ   SEQUENCE   555 AA;  61056 MW;  6B638D6E02279887 CRC64;

  Query Match             6.9%;  Score 148.5;  DB 12;  Length 555;
  Best Local Similarity   22.7%;  Pred. No. 9.9e-18;
  Matches   90;  Conservative   45;  Mismatches  139;  Indels  124;  Gaps   18;

Qy       5 PPPPVPV PVQPSPPPPAPLPA-LPPDEALAPLDJNGAPSPP--------MLGPSPGPG 55
           ||  |  |:|: |     | | |  |   :    |: :  |        : :| |
Db     190 QPAPAPTRPIAPAPPEPTRQETPIAPASEPAKQESPIGSSAPEPIRQETPTGDPAPPP--P 247

Qy      56 PPAVAAPPAPLQLGPSSSIPEPQGPLPDLPGLHIPIAHTPPPPPAPPIPLPEDP-GTL 114
           |:|    |  |||                ||      ||||| ||   |:|| | :
Db     248 PTPPPPPPPEPLQ---  ------------QKSSAVPPPPPPPLPPPGAPDDPPGEI 288

Qy     115 SAPPPLTLPVEPLGVSTQLLA------- --PSQSVCPSYTGTPWKIP--LPPEVPY--- 160
           :|  :|  :  | ||       |:     :|  ||  ||  :|: |
Db     289 QQPVR  PKPAAERAPIDALPAETRRGVQLPIAIERAPIY--TPDALPAEIPQGVPLPPA 344

Qy     161         PPENPCHPYYLPLLCEQILPDTPSEGCPIPIGLNPPPPMPMPLLGTLEVDLQPLT 213
                   |: :   |  |                 |   :: |:| :          | |
Db     345 APPAPPPPPSSPAPLLL          - -  - -PIPNPDPIPIPPVAPPAIPPPASAI 387

Qy     214 TTP  DSVKKEI    VVAAPPDWANTPPPPPVSPSPSPSPVQLIAVSHPSLPLLPVT-G 272
           : :|:    : :| :: ||| |        | || :  ::    ||:     |
Db     388 NNPLMQLLNRALRGMPMQLAEELDANYIS       SDWSDAEDSSLPPDALPIPLALIDG 440

Qy     273 PLLSPDQLPLLGSYSPAEVLSVP-PPPSTIELPLPPEQLVLHTAPAPAIF---------- 317
           :| |  :      : | :          ||| :  |    :|||||
Db     441 PRLSESERPSIAP-  PLAGAP----------IPSAPVTLRLGAPAIPIDNRILSPFY 486

Qy     318  ALVELPLNELKRDSGYVIALPSYITPNCSILSPHPG 353
            :|:|||     :: |          |:||
Db     487 QLTAPLYLHDLKLPEGAVLDL          PNRG 521


RESULT 13
Q91W51
ID   Q91W51      PRELIMINARY;      PRT;   559 AA.
AC   Q91W51;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Wasp family 1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RA   Strausberg R.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC016896; AAH16896.1;  -.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
SQ   SEQUENCE   559 AA;  61498 MW;  FAF04A64C18CC4F1 CRC64;

  Query Match             6.2%;  Score 142.5;  DB 11;  Length 559;
  Best Local Similarity   30.5%;  Pred. No. 0.0056;
  Matches   47;  Conservative   11;  Mismatches   47;  Indels   44;  Gaps    8;

Qy       1 MTQSRI---  SPVPVPVQPLPPPAPLPPLPPPPEALAPLDINGAPSP  PPMLSPSPGPG 54
           : ::||     |  || | |: ||       |   ||:  :  : :|  || : :|
Db     303 LIENRPQSPAAGRTPVPVSPTSPPPP     PPLPSALSTSSLPAPMTSTPPPPVPPPP 356

Qy      54 PSPPAVA-     PPPPAPLPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPL 112
             |||| |     : | |||: |                          : :  ::| :||
Db     357   PPPATALPAPAVPPPPAPLPIAPG                     VPPPAPPPIAPPPVP 404

Qy     108 PEPPGTLPAPPPGLTQPVEPQGVSTQLLAPSGSV 141
           |  |     |   | | ||           | | |
Db     396 PSPP-VAPAAPVPETVPVHP ---- --LPQPEV 429


RESULT 14
Q9ERQ9
ID   Q9ERQ9      PRELIMINARY;      PRT;   559 AA.
AC   Q9ERQ9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   WAVE-1.
GN   WASF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20478428; PubMed=10970852;
RA   Westphal R.S., Soderling S.H., Alto N.M., Langeberg L.K., Scott J.D.;
RT   "Scar/WAVE-1, a Wiskott-Aldrich syndrome protein, assembles an actin
RT   associated multi-kinase scaffold.";
RL   EMBO J. 19:4589-4600(2000).
DR   EMBL; AF290877; AAG02214.1;  -.
DR   MGD; MGI:1891...; Wasf1.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00246; WH2; 1.
SQ   SEQUENCE   559 AA;  61506 MW;  8746310987D80D16 CRC64;

  Query Match             6.2%;  Score 141.5;  DB 11;  Length 559;
  Best Local Similarity   30.5%;  Pred. No. 0.0066;
  Matches   47;  Conservative   11;  Mismatches   47;  Indels   44;  Gaps    8;

Qy       1 MTQSRI---  SPVPVPVQPLPPPAPLPPLPPPPEALAPLDINGAPSP  PPMLSPSPGPG 54
           : ::||     |  || | |: ||       |   ||:  :  : :|  || : :|
Db     303 LIENRPQSPAAGRTPVPVSPTSPPPP-----PPLPSALSTSSLPAPMTSTPPPPVPPPP 356
```

```
Qy      54 KGPPPAVA----- PRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPL 107
           |||| |       | | |||| | |                  : |  ||| ||||
Db     357 --PPPATALQAPAVPPPPAPLQIAPG------------------VLHPAPPPIAPPLVQ 395

Qy     108 PEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSV 141
           |  |    |   | | ||           | | |
Db     396 PSPP-VARAAPVCETVPVHP--------LPQGEV 420


RESULT 15
Q8R5H6
ID   Q8R5H6      PRELIMINARY;      PRT;   559 AA.
AC   Q8R5H6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Wave1/Scar.
GN   WAVE1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RA   Benachenhou N., Massy I., Vacher J.;
RT   "Characterization and expression analyses of the mouse WASP family
RT   member Wave1/Scar.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF467773; AAL75574.1; -.
SQ   SEQUENCE   559 AA,  61524 MW;  7178BC4987DCB157 CRC64;

  Query Match              6.2%;  Score 141.5;  DB 11;  Length 559;
  Best Local Similarity   30.5%;  Pred. No. 0.0066;
  Matches   47;  Conservative   11;  Mismatches   47;  Indels   49;  Gaps    8;

Qy       1 MYQSRP-----GPVPVPVQLSRPPKAFLPKIDPKDPALAKLGINGAHSS--PPMLSPSPG 53
             : ::||     |  || | |: ||        |   ||:   :  : :|  || : | |
Db     303 LIENRPQSPAAGRTPVFVSPTPPPPP-----PPLPSALSTSSLRASMTSTPPPPVPPPP- 356

Qy      54 KGPPPAVA---  PRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPL 107
           |||| |      | | ||||: |                  : |  ||| ||||
Db     357 --PPPATALQAPAVPPPPAPLQIAPG------------------VLHPAPPPIAPPLVQ 395

Qy     108 PEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSV 141
           |  |    |   | | ||           | | |
Db     396 PSPP-VARAAPVCETVPVHP--------LPQGEV 420


Search completed: January 10, 2003, 02:22:41
Job time : 39.3367 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein protein search, using sw model

Run on: January 10, 2003, 02:19:59 ; Search time 14.3391 Seconds
(without alignments)
896.699 Million cell updates/sec

Title: US-09-803-126-4
Perfect score: 2287
Sequence: 1 MYQSREPGFVIVVPQLSRIFP ... WDPASEVPKMGPGQAEARPA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	134	5.9	525	4	US-08-764-870-7	Sequence 7, Appli
2	134	5.9	525	4	US-08-980-115-7	Sequence 7, Appli
3	133	5.8	533	1	US-07-952-800-2	Sequence 2, Appli
4	119	5.2	538	4	US-09-309-572-12	Sequence 12, Appl
5	119	5.2	1737	4	US-09-309-572-13	Sequence 13, Appl
6	118.5	5.2	174	4	US-08-818-112-143	Sequence 143, App
7	118.5	5.2	174	4	US-08-818-111-138	Sequence 138, App
8	118.5	5.2	174	4	US-09-056-556-143	Sequence 143, App
9	118.5	5.2	174	4	US-09-072-596-138	Sequence 138, App
10	118	5.2	214	1	US-08-217-327-4	Sequence 4, Appli
11	118	5.2	1248	2	US-09-080-897-2	Sequence 2, Appli
12	118	5.2	1248	4	US-09-323-735-2	Sequence 2, Appli
13	117.5	5.1	1162	2	US-08-728-322A-2	Sequence 2, Appli
14	117.5	5.1	1162	4	US-09-228-568-2	Sequence 2, Appli
15	117	5.1	538	4	US-09-370-368-9	Sequence 3, Appli
16	117	5.1	1315	3	US-08-899-595-3	Sequence 4, Appli
17	116	5.1	446	1	US-07-952-800-4	Sequence 4, Appli
18	115.5	5.1	1255	2	US-09-080-897-4	Sequence 1, Appli
19	115.5	5.1	1255	3	US-08-899-595-1	Sequence 4, Appli
20	115.5	5.1	1255	4	US-09-323-735-4	Sequence 40, Appl
21	114.5	5.0	248	4	US-08-630-915A-40	Sequence 194, App
22	114.5	5.0	565	4	US-08-630-915A-194	Sequence 2, Appli
23	114	5.0	1274	4	US-09-095-443-2	Sequence 16, Appl
24	113.5	5.0	267	3	US-08-301-162-16	Sequence 16, Appl
25	113.5	5.0	267	4	US-09-461-240-16	Sequence 16, Appl
26	113.5	5.0	267	4	US-09-968-927-16	Sequence 4, Appli
27	113.5	5.0	380	2	US-09-026-587-4	Sequence 4, Appli
28	113.5	5.0	380	2	US-09-227-420-4	Sequence 6, Appli
29	113	4.9	902	1	US-08-306-479B-6	Sequence 6, Appli
30	113	4.9	902	1	US-08-818-823-6	Sequence 2, Appli
31	113	4.9	983	4	US-09-311-236-2	Sequence 2, Appli
32	112.5	4.9	2414	1	US-08-227-536-2	Sequence 2, Appli
33	112.5	4.9	2414	5	PCT-US95-04682-2	Sequence 10, Appl
34	112	4.9	1291	4	US-09-150-460B-10	Sequence 6, Appli
35	112	4.9	1291	4	US-09-220-641-5	Sequence 54, Appl
36	111.5	4.9	398	4	US-09-303-064-54	Sequence 54, Appl
37	111.5	4.9	398	4	US-09-086-503-54	Sequence 2, Appli
38	111.5	4.9	2972	4	US-09-579-181-2	Sequence 1, Appli
39	111.5	4.9	3118	4	US-09-579-181-1	Sequence 6, Appli
40	110	4.8	306	1	US-08-217-327-6	Sequence 9, Appli
41	107.5	4.7	905	2	US-08-574-959A-9	Sequence 9, Appli
42	107.5	4.7	905	4	US-09-357-014-9	Sequence 7, Appli
43	107.5	4.7	1135	2	US-08-574-959A-7	Sequence 7, Appli
44	107.5	4.7	1135	4	US-09-357-014-7	Sequence 6, Appli
45	105.5	4.6	297	2	US-08-580-545B-6	

ALIGNMENTS

RESULT 1
US-08-764-870-7
; Sequence 7, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletterick, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Pushner, Peter J
; APPLICANT: Apriletti, James W
; APPLICANT: West, Brian
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
; TITLE OF INVENTION: Binding Domains
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,540
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,543
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,606
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Jackie N
; REGISTRATION NUMBER: 35,966
; REFERENCE/DOCKET NUMBER: UCAL-246/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid

```
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-764-877-7

  Query Match             5.9%;  Score 134;  DB 4;  Length 525;
  Best Local Similarity   22.7%;  Pred. No. 0.00037;
  Matches   93;  Conservative   46;  Mismatches  155;  Indels  116;  Gaps   19;

Qy       4 SPPGPVPVPVQPSRPPEAFLRKIDPKDEALAKLGINGAHSSPPMLSP------------ 50
           | |:|  | |  :|   :    |   :  :|| :||   |||  |
Db      78 SSPNPLPQGVPPPSPPGPPL----PPSTA-PTLGGSGAPPPPPMPPPPLGSPFPVISSSM 132

Qy      51  SPGPGPP- ---- PAVAPPPPAPLQL-----GPSSSIKEKQGPLLPIPGQPLPIAHTP 97
             :|: ||       |  :|:  : : |     ||   :|    |:| : |       | |
Db     133 GSPGLPPPAPPGPSGPVSSPQINSTVSLPGGSNGPPPIVPP   PVLGVPG-----LHCP 184

Qy      98 PPPPAPPLPLPEDPGTLSAERPCLTGPVEDQGVSPQLLAPSGSVCPSYTGTPWKLPLRKE 157
           |:|  |          : :| | |     |   : :  |   |       :|  :||:
Db     185 PPPGGPG----------AGKPLCAICGDRSSGKHYGVYSCEGCKGF------FKRTIRKD 228

Qy     158 VPY              PPPNPGHPVYYIPLICPQIIPGTTPSPSCIRISQNPPPPMPT---- 200
           :                |   : | :  :|:   |       : :| | ||
Db     229 LTYSCRDNKDCIVDKRQRNRCQYCRYQKCLATGMKREAVQE-----EPQPGPIPGDGE 282

Qy     201 LLGGL -EVPLDSLTTTEDSVKKPIVVAARDNWANYFSPPPPVSGESGS --- ------ 247
           :    |: :|  :   | :|:::                          :|:|||
Db     283 CAGGAPEEMPVDRILEAELAVEQKSDQGVEGPGG       TGSGSGSPNDPVTNICQ 333

Qy     248 -DVQLLAVSHKGLKLLKVTQQPGLKPTQLKIIGNYSEPAEVL--------GVEWPGGSTLE 298
           | ||  :     |:   :  |    |   ||    : |:|         :: | | |
Db     334 AADKQLFTLVEWAKRIPHFSSLP--LDDQVILLRAGWNELLIASFSHRSIDVRDGILLA 390

Qy     299 LSLKSEQLVLHTARRAI  EALVELF  LNELKRDNGYVIALRGYTITDN 344
           |   :  |:|   ||    : ||   : ::: |   :  :|: | |
Db     391 TGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMRMDKTELGCLRAIILFN 440
```

```
RESULT 2
US-08-980-115-7
; Sequence 7, Application US/08980115
; Patent No. 6266622
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Thomas S.
;  APPLICANT: Baxter, John D.
;  APPLICANT: Fletterick, Robert J.
;  APPLICANT: Wagner, Richard L.
;  APPLICANT: Kushner, Peter J.
;  APPLICANT: Apriletti, James W.
;  APPLICANT: West, Brian L.
;  APPLICANT: Shiau, Andrew K.
;  TITLE OF INVENTION: NUCLEAR RECEPTOR LIGAND ASSAY AND LIGAND BINDING DOMAIN
;  FILE REFERENCE: UCAL-246/02US
;  CURRENT APPLICATION NUMBER: US/08/980,115
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: 08/764,870
;  EARLIER FILING DATE: 1996-12-16
;  EARLIER APPLICATION NUMBER: 60/[illegible]
;  EARLIER FILING DATE: 1995-12-14
;  EARLIER APPLICATION NUMBER: 60/008,543
;  EARLIER FILING DATE: 1995-12-13
;  EARLIER APPLICATION NUMBER: 60/008,540
;  EARLIER FILING DATE: 1995-12-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 7
;  LENGTH: 525
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: DOMAIN
;  LOCATION: (292)..(523)
;  OTHER INFORMATION: minimal ligand binding domain
```

```
US-08-980-115-7

  Query Match             5.9%;  Score 134;  DB 4;  Length 525;
  Best Local Similarity   22.7%;  Pred. No. 0.00037;
  Matches   93;  Conservative   46;  Mismatches  155;  Indels  116;  Gaps   19;

Qy       4 SPPGPVPVPVQPSRPPEAFLRKIDPKDEALAKLGINGAHSSPPMLSP------------ 50
           | |:|  | |  :|   :    |   :  :|| :||   |||  |
Db      78 SSPNPLPQGVPPPSPPGPPL----PPSTA-PTLGGSGAPPPPPMPPPPLGSPFPVISSSM 132

Qy      51 -SPGPGPP-------PAVAPPPPAPLQL----GPSSSIKEKQGPLLPIPGQPLPIAHTP 97
            |||  ||       |  :|:  : : |     ||   :|    |:| : |       | |
Db     133 GSPGLPPPAPPGPSGPVSSPQINSTVSLPGGSNGPPPIVPP   PVLGVPG-----LHCP 184

Qy      98 PPPPAPPLPLPEDPGTLSAERPCLTGPVEDQGVSPQLLAPSGSVCPSYTGTPWKLPLRKE 157
           |:|  |          : :| | |     |   : :  |   |       :|  :||:
Db     185 PPPGGPG----------AGKPLCAICGDRSSGKHYGVYSCEGCKGF------FKRTIRKD 228

Qy     158 VPY-------------PPNPGHPVYYIPLICPQIIPGTTPSPSCIRISQNPPPPMPT---- 200
           : |               |   : | :  :|:   |       : :| | ||
Db     229 LTYSCRDNKDCIVDKRQRNRCQYCRYQKCLATGMKREAVQE-----EPQPGPIPGDGE 282

Qy     201 LLGGL -EVPLDSLTTTEDSVKKPIVVAARDNWANYFSPPPPVSGESGS             247
           |    |: :|  :   | :|:::                          :|:|||
Db     283 CAGGAPEEMPVDRILEAELAVEQKSDQGVEGPGG       TGSGSGSPNDPVTNICQ 333

Qy     248 --DVQLLAVSHKGLKLLKVTQQPGLKPTQLKIIGNYSEPAEVL--------GVEWPGGSTLE 298
             | ||  :     |:   :  |    |   ||    : |:|         :: | | |
Db     334 AADKQLFTLVEWAKRIPHFSSLP--LDDQVILLRAGWNELLIASFSHRSIDVRDGILLA 390

Qy     299 LSLKSEQLVLHTARRAI  EALVELF  LNELKRDNGYVIALRGYTITDN 344
           |   :  |:|   ||    : ||   : ::: |   :  :|: | |
Db     391 TGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMRMDKTELGCLRAIILFN 440
```

```
RESULT 3
US-07-952-800-2
; Sequence 2, Application US/07952800
; Patent No. 5403925
;  GENERAL INFORMATION:
;    APPLICANT:  OZATO, KEIKO
;    TITLE OF INVENTION:  A NEW MEMBER OF THE NUCLEAR HORMONE
;    TITLE OF INVENTION:  RECEPTOR SUPERFAMILY AND A DNA [illegible]
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND AND TOWNSEND
;      STREET:  1 MARKET PLAZA, STEUART TOWER, SUITE [illegible]
;      CITY:  SAN FRANCISCO
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/952,800
;      FILING DATE:  19920928
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  15290 21 1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415-543-5043
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  533 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-07-952-800-2

Query Match 5.8%; Score 133; DB 1; Length 533;
Best Local Similarity 22.8%; Pred. No. 0.00034;
Matches 93; Conservative 46; Mismatches 153; Indels 120; Gaps 19;

Qy 4 SPPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSP---------- 50
Db 86 SSPNPLPQGVPPPSPPGPPL----PPSTA-PSLGGSGAPPPPPMPPPPLGSPFPVISSSM 140
Qy 51 SPGKGPP-------PAVAPRPKAPLQL------GPSSSIKEKQGPILDLFGQKLPIAHTP 97
Db 141 GSPGLFPPAPPGPSGPVSSPLINSTVSLPGGSGPFPEDVPP---PVIAGVRG-----IHTP 192
Qy 98 PPPPAPPIPLPEDPGTLSAEPPCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKE 157
Db 193 PPPGGPG----------AGFPLPAIGGPFSGGPHYGVYSCEGCKGF------FFFTIPYD 236
Qy 158 VPY-------------PRENFSHPYYLRLLCEQILRDTFSESCIRISQNERPKMKDL--- 201
Db 237 LTYSCRDNKDTTVDEFCRNRCQYCFYQPCLATGMKREAV--------QEERQPGSKDKDGD 288
Qy 202 ---LGGL--EVDLDSLTTTEDSVKKRIVVAARDNWANYFSPFFPVSGESGS--------- 247
Db 289 GEGAGGAPEEMPVDRILEAELAVEQKSDQGVEGPGG---------TGGSGSSPNDPVTNI 339
Qy 248 ----DVQLLAVSHRGLRLLKVTQGFGLFPDQLKILQSYSKAEVL--- -GVECRGGST 296
Db 340 CQAADKQLFTLVEWAKRIPHFSSLP---LDDQVILLRAGWNELLIASFSHRSIDVRDGIL 396
Qy 297 LELSLKGEQIVLHTARARAI--EALVELF--LNELKKDSGYVIALRSYITDN 344
Db 397 LATGLHVHRNSAHSAGVGAIFDRVLTELVSKMPDMRMDKTELGCLRAIILFN 448

RESULT 4
US-09-309-572-12
; Sequence 12, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: gag-prorein
US-09-309-572-12

Query Match 5.2%; Score 119; DB 4; Length 538;
Best Local Similarity 23.8%; Pred. No. 0.0072;
Matches 63; Conservative 32; Mismatches 98; Indels 72; Gaps 14;

Qy 6 PGPVPVP-VQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRP 64
Db 93 PPPWVKPFVHPKPPP--------PLPPSAPSLPLEPPPRSTPPRSSLYPALTPSLGAKPKP 144
Qy 65 KAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPP---------APPLPLPEDPGTLS 115
Db 145 QV---------LSDSGGPLIDLLTEDPPPYRDPRPPPSDRDGNGGEATPAGEAPDPSPMA 195
Qy 116 AEPPCLTQPVEDQGVSTQL- LAPSGSVCFSYTGTPWKLFLRKEVPYPREN---FSH-PY 169
Db 196 SRLRGRREPPVADSTTSQAFPLRAGGNGQLQY----WP-FSSSDLYNWKNNPSFSEDPG 250
Qy 170 YLRLLCEQIL---RDTFSESCIRISQNERPKMKDLLGGLEVDLDSLTTTEDSVKKRIVVA 226
Db 251 KLTALIESVLITHQPTWDD----------CQQLLG-------TLLTGEE--KQRVLLE 289
Qy 227 APDNWANYFSPFFPVSGESGSDVQL 251
Db 290 APK----------AVRGDDGRPTQL 304

RESULT 5
US-09-309-572-13
; Sequence 13, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1737
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: gag-pol protein
US-09-309-572-13

Query Match 5.2%; Score 119; DB 4; Length 1737;
Best Local Similarity 23.8%; Pred. No. 0.04;
Matches 63; Conservative 32; Mismatches 98; Indels 72; Gaps 14;

Qy 6 PGPVPVP-VQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRP 64
Db 93 PPPWVKPFVHPKPPP--------PLPPSAPSLPLEPPPRSTPPRSSLYPALTPSLGAKPKP 144
Qy 65 KAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPP---------APPLPLPEDPGTLS 115
Db 145 QV---------LSDSGGPLIDLLTEDPPPYRDPRPPPSDRDGNGGEATPAGEAPDPSPMA 195
Qy 116 AEPPCLTQPVEDQGVSTQL- LAPSGSVCFSYTGTPWKLFLRKEVPYPREN---FSH-PY 169
Db 196 SRLPGPPEPPVADSTTSQAFPLRAGGNGQLQY----WP-FSSSDLYNWKNNPSFSEDPG 250
Qy 170 YLRLLCEQIL---RDTFSESCIRISQNERPKMKDLLGGLEVDLDSLTTTEDSVKKRIVVA 226
Db 251 KLTALIESVLITHQPTWDD---------CQQLLG-------TLLTGEE--KQRVLLE 289
Qy 227 APDNWANYFSRFFPVSGESGSDVQL 251
Db 290 APK----------AVRGDDGRPTQL 304

RESULT 6
US-08-818-112-143
; Sequence 143, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION.
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

```
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104 7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/818,112
;       FILING DATE:  13-MAR-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.41106
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  143:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  174 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-818-112-143

  Query Match              5.2%;  Score 118.5;  DB 4;  Length 174;
  Best Local Similarity   28.2%;  Pred. No. 0.0016;
  Matches   52;  Conservative    9;  Mismatches   47;  Indels   71;  Gaps    5;

Qy          5 RPGPVPVPVQPSRP----------PEAPLRKIDPKDPALAKLGINGAHSSPPEMLSPSPGKIP 56
                ||    |||   |    |          |:|  |                     ||  |:  :|     |
Db         18 RPPVPPVPPLPPSPPSPPTGWVPRALL---------------------PPWLAGTPPAPP 56

Qy         57 PPAVAP-RPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTP------------------P 98
                   |  :||    |  |||                ||   |      ||  :|  |                  |
Db         57 VPPMAPLPPAAPLP-------------PLPPL--PPLPTSHPPPPPAPPAPPAPPACPFVP 102

Qy         99 PPPAPPLPLPPEDPGTLSAERRC-------LTQPVEDQGVSTQLLAPSGSVCFSYTG 147
                  ||||||   |  |   |  |:   |          |  |    |  |  :  |  :|:   |  |
Db        103 VPPAPPLP PSPPTELPADAACPPAPPAPPLAPPSPPAG-SAAIPALTGATSASTLG 157

RESULT 7
US-08-818-111-138
; Sequence 138, Application US/08818111
; Patent No. 6338852
;   GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G.
;     APPLICANT:  Skeiky, Yasir A.W.
;     APPLICANT:  Dillon, Davin C.
;     APPLICANT:  Campos-Neto, Antonia
;     APPLICANT:  Houghton, Raymond
;     APPLICANT:  Vedvick, Thomas S.
;     APPLICANT:  Twardzik, Daniel R.
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;     NUMBER OF SEQUENCES:  148
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104 7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/818,111
;       FILING DATE:  13-MAR-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.41706
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  138:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  174 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-818-111-138

  Query Match              5.2%;  Score 118.5;  DB 4;  Length 174;
  Best Local Similarity   28.2%;  Pred. No. 0.0016;
  Matches   52;  Conservative    9;  Mismatches   47;  Indels   71;  Gaps    5;

Qy          5 RPGPVPVPVQPSRP----------PEAPLRKIDPKDPALAKLGINGAHSSPPEMLSPSPGKIP 56
                ||    |||   |    |          |:|  |                     ||  |:  :|     |
Db         18 RPPVPPVPPLPPSPPSPPTGWVPRALL---------------------PPWLAGTPPAPP 56

Qy         57 PPAVAP-RPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTP------------------P 98
                   |  :||    |  |||                ||   |      ||  :|  |                  |
Db         57 VPPMAPLPPAAPLP-------------PLPPL--PPLPTSHPPPPPAPPAPPAPPACPFVP 102

Qy         99 PPPAPPLPLPPEDPGTLSAERRC-------LTQPVEDQGVSTQLLAPSGSVCFSYTG 147
                  ||||||   |  |   |  |:   |          |  |    |  |  :  |  :|:   |  |
Db        103 VPPAPPLP PSPPTELPADAACPPAPPAPPLAPPSPPAG-SAAIPALTGATSASTLG 157

RESULT 8
US-09-056-556-143
; Sequence 143, Application US/09056556
; Patent No. 6350456
;   GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G.
;     APPLICANT:  Skeiky, Yasir A.W.
;     APPLICANT:  Dillon, Davin C.
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND     TRE
;     NUMBER OF SEQUENCES:  241
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104 7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/056,556
;       FILING DATE:  07-APR-1998
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.457
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  143:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  174 amino acids
```

```
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
S-09-056-556-143

 Query Match                 5.2%;  Score 118.5;  DB 4;  Length 174;
 Best Local Similarity  28.2%;  Pred. No. 0.0016;
 Matches   50;  Conservative    9;  Mismatches   47;  Indels   71;  Gaps    9;

Qy       5 RPGPVPVPVQPSRP--------PKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGP 56
           ||   |||  |  |        |:| |                  || |: :|    |
Db      18 RPPVPPVPPLPPSPPSPPTGWVPRALL--------------------PPWLAGTPPAPP 56

Qy      57 PPAVAP-RPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTP----------------P 98
           | :||  | |||                 || |   || :| |                 |
Db      57 VPPMAPLPPAAPLP------------PLPPL--PPLPTSHPPRPPAPPAPPAPPACPFVP 102

Qy      99 PPPAPPLPLPEDPGTLSAEPPC---------LTQPVEDQGVSTQLLAPSGSVCFSYTG 147
           |||||||  |  |  | |  |:         |  | | | . | :|.  |  |
Db     103 VPPAPPLP-PSPPTELPADAACPPAPPAPPLAPPSPPAG-SAAIRALTGATSASTLG 157


RESULT 9
US-09-072-596-138
; Sequence 138, Application US/09072596
; Patent No. 6458366
  GENERAL INFORMATION:
;   APPLICANT:  Reed, Steven G.
;   APPLICANT:  Skeiky, Yasir A.W.
;   APPLICANT:  Dillon, Davin C.
;   APPLICANT:  Campos Neto, Antonia
;   APPLICANT:  Houghton, Raymond
;   APPLICANT:  Vedvick, Thomas S.
;   APPLICANT:  Twardzik, Daniel R.
;   APPLICANT:  Lodes, Michael J.
;   APPLICANT:  Hendrickson, Ronald C.
;   TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;   NUMBER OF SEQUENCES:  350
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  SEED and BERRY LLP
;     STREET:  6300 Columbia Center, 701 Fifth Avenue
;     CITY:  Seattle
;     STATE:  Washington
;     COUNTRY:  USA
;     ZIP:  98104-7092
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/072,596
;     FILING DATE:  05-MAY-1998
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Maki, David J.
;     REGISTRATION NUMBER:  31,392
;     REFERENCE/DOCKET NUMBER:  210121.417C9
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (206) 622-4900
;     TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  138:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  174 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
US-09-072-596-138

 Query Match                 5.2%;  Score 118.5;  DB 4;  Length 174;
 Best Local Similarity  28.2%;  Pred. No. 0.0016;
 Matches   50;  Conservative    9;  Mismatches   47;  Indels   71;  Gaps    9;

Qy       5 RPGPVPVPVQPSRP--------PKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGP 56
           ||   |||  |  |        |:| |                  || |: :|    |
Db      18 RPPVPPVPPLPPSPPSPPTGWVPRALL--------------------PPWLAGTPPAPP 56

Qy      57 PPAVAP-RPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTP----------------P 98
           | :||  | |||                 || |   || :| |                 |
Db      57 VPPMAPLPPAAPLP------------PLPPL--PPLPTSHPPRPPAPPAPPAPPACPFVP 102

Qy      99 PPPAPPLPLPEDPGTLSAEPPC---------LTQPVEDQGVSTQLLAPSGSVCFSYTG 147
           |||||||  |  |  | |  |:         |  | | | . | :|.  |  |
Db     103 VPPAPPLP-PSPPTELPADAACPPAPPAPPLAPPSPPAG-SAAIRALTGATSASTLG 157


RESULT 10
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
  GENERAL INFORMATION:
;   APPLICANT:  John, Maliyakal E
;   APPLICANT:  Barton, Kenneth A
;   TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;   NUMBER OF SEQUENCES:  16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Quarles and Brady
;     STREET:  P.O. Box 2113
;     CITY:  Madison
;     STATE:  WI
;     COUNTRY:  USA
;     ZIP:  53701-2113
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/217,327
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/812,233
;     FILING DATE:  19-DEC-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Seay, Nicholas J
;     REGISTRATION NUMBER:  27,386
;     REFERENCE/DOCKET NUMBER:  1122990831
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  608-251-5000
;     TELEFAX:  608-251-9166
;   INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  214 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-217-327-4

 Query Match                 5.2%;  Score 118;  DB 1;  Length 214;
 Best Local Similarity  27.9%;  Pred. No. 0.0024;
 Matches   38;  Conservative    7;  Mismatches   57;  Indels   34;  Gaps    4;

Qy       6 PGPVPVPVQPSPPPKAPLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPAA    61
           | ||  |   | ||                  :||| :|  |  ||||    :
Db      58 PPPVSTPPPTSSPPP--------------------VTASPPPVSTPPPSSPPPAIPPAS 97

Qy      62 PPPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAEPPCL 121
           | | |  |  |   :          |  |  ||||  || ||   | |: |
Db      98 PPPATPPPASPPPATPPPASPPPATPPPATPPPATPPPATPPPAPLASPPATVPA 152

Qy     122 TQPVEDQGVSTQLLAP 137
           ||:    | | :|
Db     153 ISPVQ-----TPLTSP 163
```

```
RESULT 11
US-09-080-897-2
; Sequence 2, Application US/09080897
; Patent No. 6086874
;  GENERAL INFORMATION:
;    APPLICANT:  King, Mary-Claire
;    APPLICANT:  Lynch, Eric D.
;    APPLICANT:  Lee, Ming
;    APPLICANT:  Morrow, Jan E.
;    APPLICANT:  Welcsh, Piri L.
;    APPLICANT:  Leon, Pedro E.
;    TITLE OF INVENTION:  Modulators of Actin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  [illegible] & [illegible] LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/080,897
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UW97-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1248 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-080-897-2

  Query Match             5.2%;  Score 118;  DB 2;  Length 1248;
  Best Local Similarity   23.6%;  Pred. No. 0.021;
  Matches   95;  Conservative   40;  Mismatches  154;  Indels  114;  Gaps   20;

Qy       4 SRPGPVPVPVQPSRPPKAFLPKIIPKPAPALAPIGINGAHSSPPMLSPSPGKGPPP----- 58
           : ...  :: :  | |  |   :| |                  || |  | |  |||
Db     647 SLPGGTAIPPPPPLPGSA---RIPP---------------PPPPLPGSAGIPPPPPPLPG 688

Qy      59  AVAEPPEAELQLGLHVRGDEIHFRPGFKRLEIAHGTPPPPPAPPIPLPEDPGTLSA 116
            |  :  :  ||  ||                          |||| |  |  :|  |  :
Db     689 EAGMPFPPPPLPGGPG----------  ---------IPPPPPPPGGPGIPPPPPGMGM 726

Qy     117 RRRCLTQPVEDQGVSTQLLAPGSSVCPSYTGTPWKLPLRKEVEYPPRENFSHPYYLPLLCE 176
                 |  ||    : | |         | ||:: : ||   | ||:|     :|: |
Db     727 P---  PPPPPGPGVPAAPVLPPGL       TPPKLY-PPEVQLPPPNWS-----KLVAE 769

Qy     177 QILRDTPSESGIPISQNERPPMPINTGPRVRDSITTT ------- PDSVKYPIVVAA 227
           :  :: :         : :|  :  :|   | :   : | |           : ||:|: |
Db     770 LLSQDCPWTKV    REEPPENNELPAPLTLTPGAQTPTPPL-PQTSPPPVQLPPPVPPI 825

Qy     228 R   DNWANYPSRF   PPVSGESGSDVQLLAVSHPGL------PIIFVTGQPGIPPTQL 275
           :     :  | |  :   :| :| :  :    |      |:| | | |   |.||
Db     826 EVLPSRLAQNLPIPLGSPPMPYQEIPNV ILPVNPAVITPPMIQNLIP--QMP--PPPQL 880

Qy     276 RILP  ----  ---SYSPAEVLGVECRGGSTLELSLKSEQLVLHTAPAPAIEALVELPL 324
             ::          | | |:| | | |  |  |     :    |
Db     881 PMLPLRGNVTPPLAESEPSPTTMQTTPELSTPELNAIIPRLPPRPQVENIPPEIVSVTAAT 940

Qy     325 NELPRESGY---  VIALPSRI  -IDNGSLLSPHRGSGIKL 366
             ||:|   :       : : :|:   : |      |:    |
Db     941 EELRKSESPSNLLEITLLVGNTMNAGRNAGAPGPNISPLTEL 984


RESULT 12
US-09-323-735-2
; Sequence 2, Application US/09323735
; Patent No. 6197932
;  GENERAL INFORMATION:
;    APPLICANT:  King, Mary-Claire
;    APPLICANT:  Lynch, Eric D.
;    APPLICANT:  Lee, Ming
;    APPLICANT:  Morrow, Jan E.
;    APPLICANT:  Welcsh, Piri L.
;    APPLICANT:  Leon, Pedro E.
;    TITLE OF INVENTION:  Modulators of Actin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/323,735
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/080,897
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UW97-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1248 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-323-735-2

  Query Match             5.2%;  Score 118;  DB 4;  Length 1248;
  Best Local Similarity   23.6%;  Pred. No. 0.021;
  Matches   95;  Conservative   40;  Mismatches  154;  Indels  114;  Gaps   20;

Qy       4 SRPGPVPVPVQPSRPPKAFLPKIIPKPAPALAPIGINGAHSSPPMLSPSPGKGPPP----- 58
           | ||    :|  |  |  |   :| |                  :|  :    :  |||
Db     647 SLPGGTAIPPPPPLPGSA---RIPP----------      PPPPLPGSAGIPPPPPPLPG 688

Qy      59  AVAEPPEAELQLGLHVRGDEIHFRPGFKRLEIAHGTPPPPPAPPIPLPEDPGTLSA 116
            | | |  :|  ||                          |||| |  |  :|  |  :
Db     689 EAGMPFPPPPLPGGPG                          IPPPPPPPGGPGIPPPPPGMGM 726

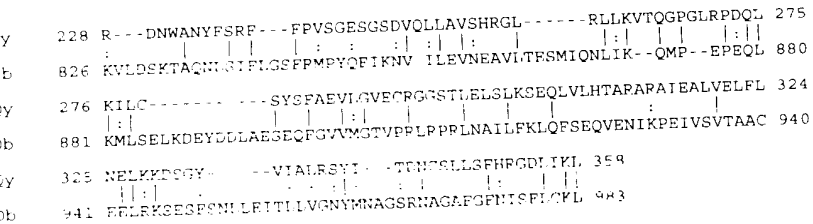
Qy     117 RRRCLTQPVEDQGVSIQLLAPGSSVCPSYTGTPWKLPLRKEVEYPPRENFSHPYYLPLLCE 176
                 |  ||    : | |         || ||: : ||   | ||:|     :|: |
Db     727 P---  PPPPPGPGVPAAPVLPPGL       TPPKLY PPEVQLPPPNWS     KLVAE 769

Qy     177 QILRDTPSESGIPISQNERPPMPINTGPRVRDSITTT          PDSVKYPIVVAA 227
           :  :: :         : :|  :  :|   | :   : | |           : ||:|: |
Db     770 LLSQDCPWTKV    REEPPENNELPAPLTLTPGAQTPTPPL-PQTSPPPVQLPPPVPPI 825

Qy     228 R   DNWANYPSRF   PPVSGESGSDVQLLAVSHPGL------PIIFVTGQPGIPPTQL 275
           :     :  | |  :   :| :| :  :    |      |:| | | |   |.||
Db     826 EVLPSRLAQNLPIPLGSPPMPYQEIPNV ILPVNPAVITPPMIQNLIP--QMP--PPPQL 880

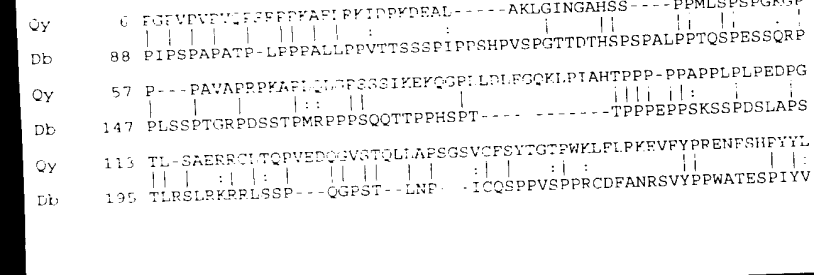
Qy     276 RILP  ----  ---SYSPAEVLGVECRGGSTLELSLKSEQLVLHTAPAPAIEALVELPL 324
             ::          | | |:| | | |  |  |     :    |
Db     881 PMLPLRGNVTPPLAESEPSPTTMQTTPELSTPELNAIIPRLPPRPQVENIPPEIVSVTAAT 940

Qy     325 NELPRESGY---  VIALPSRI  -IDNGSLLSPHRGSGIKL 366
             ||:|   :       : : :|:   : |      |:    |
Db     941 EELRKSESPSNLLEITLLVGNTMNAGRNAGAPGPNISPLTEL 984
```



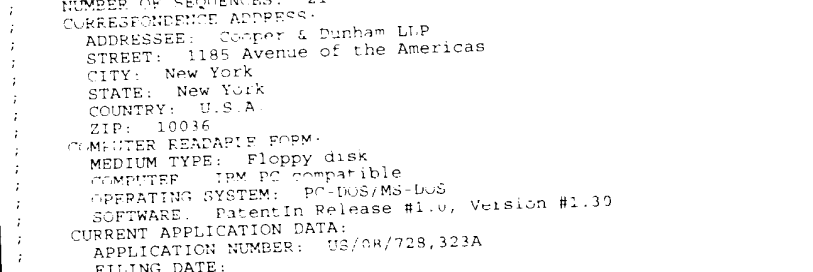
RESULT 13
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 5.1%; Score 117.5; DB 2; Length 1162;
Best Local Similarity 28.3%; Pred. No. 0.031;
Matches 51; Conservative 12; Mismatches 82; Indels 35; Gaps 10;



RESULT 14
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: PHAGING VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION. VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001P
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM. Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

Query Match 5.1%; Score 117.5; DB 4; Length 1162;
Best Local Similarity 28.3%; Pred. No. 0.031;
Matches 51; Conservative 12; Mismatches 82; Indels 35; Gaps 10;



RESULT 15
US-09-370-368-9
; Sequence 9, Application US/09370368
; Patent No. 6258932
; GENERAL INFORMATION:
; APPLICANT: Anders Vahlne
; TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.003A
; CURRENT APPLICATION NUMBER: US/09/370,368
; CURRENT FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Moloney Murine Leukemia Virus
US-09-370-368-9

Query Match 5.1%; Score 117; DB 4; Length 538;
Best Local Similarity 24.7%; Pred. No. 0.011;
Matches 62; Conservative 28; Mismatches 85; Indels 76; Gaps 14;

```
Qy      130 VSTQL  LAPSGSVCPSYTGTPWKLFLKKEVFYPREN---FSH-PYYLRLLCEQIL---R 180
            ::     |   |:    |    :  |   :::  : :    ||  |  |  |  | :|   :
Db      212 TTSQAFTLKTGSNGQLQY  - WF TGSGDLYNWKNNNPSFSEDPGKLTALIECVLITHQ 264

Qy      181 NTPSPSCIRISQNERRKMKDLLCGLEVDLDSLTTTEDSVYKPIVVAARDNWANYPSPPPP 240
            :: :              : :||,       :| | |:  |:|::: ||
Db      265 ETWGD            CQQLLG    -   TLLTGEE--KQPVLLEARK----------A 293

Qy      241 VSGESGSDVQL 251
            | :: |   ||
Db      294 VRGDDGKPTQL 304


Search completed: January 10, 2003, 02:25:37
Job time : 1e.4391 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

SM protein protein search, using sw model

Run on: January 10, 2003, 02:23:04 ; Search time 33.7992 Seconds
(without alignments)
250.842 Million cell updates/sec

Title: US-09-803-126-4
Perfect score: 2287
Sequence: 1 MYQSRPGPVPVPVQPSRPPKAF...WDRASEVRYMGPGQAEAPPA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2287	100.0	437	10	US-09-803-126-4	Sequence 4, Appli
2	2217.5	97.0	786	10	US-09-803-126-6	Sequence 6, Appli
3	2120	92.7	405	10	US-09-803-126-13	Sequence 13, Appl
4	1725	75.4	333	9	US-09-764-868-758	Sequence 758, App
5	1543.5	67.5	1337	10	US-09-803-126-1	Sequence 1, Appli
6	1537	67.2	407	10	US-09-803-126-12	Sequence 12, Appl
7	306	13.4	56	10	US-09-864-761-40685	Sequence 40685, A
8	143	6.3	559	10	US-09-858-155A-2	Sequence 2, Appli
9	131.5	5.7	1134	9	US-10-001-873-50	Sequence 50, Appl
10	129	5.6	446	10	US-09-853-386-69	Sequence 69, Appl
11	129	5.6	643	10	US-09-853-386-70	Sequence 70, Appl
12	128.5	5.6	802	10	US-09-823-240-2	Sequence 2, Appli
13	124.5	5.4	507	9	US-10-078-547-24	Sequence 24, Appl
14	124.5	5.4	625	10	US-09-853-386-63	Sequence 63, Appl
15	124.5	5.4	626	10	US-09-853-386-64	Sequence 64, Appl
16	124.5	5.4	626	10	US-09-853-386-65	Sequence 65, Appl
17	124.5	5.4	626	10	US-09-853-386-96	Sequence 96, Appl
18	124.5	5.4	731	9	US-10-086-464-17	Sequence 17, Appl
19	124	5.4	21	10	US-09-803-126-8	Sequence 8, Appli
20	123	5.4	684	10	US-09-823-240-9	Sequence 9, Appli
21	121	5.3	647	10	US-09-925-299-1002	Sequence 1002, Ap
22	120	5.2	503	9	US-10-078-547-2	Sequence 2, Appli
23	119.5	5.2	200	10	US-09-925-297-700	Sequence 700, App
24	119	5.2	463	9	US-10-029-180-80	Sequence 80, Appl
25	118.5	5.2	1567	10	US-09-835-232-2	Sequence 2, Appli
26	118	5.2	538	9	US-09-976-740-43	Sequence 43, Appl
27	118	5.2	538	12	US-10-023-529-43	Sequence 43, Appl
28	118	5.2	538	12	US-10-023-523-43	Sequence 43, Appl
29	117	5.1	538	10	US-09-827-822-9	Sequence 9, Appli
30	116	5.1	116	10	US-09-864-761-40290	Sequence [illegible], A
31	116	5.1	726	10	US-09-770-689A-4	Sequence 4, Appli
32	116	5.1	803	10	US-09-770-689A-2	Sequence 2, Appli
33	116	5.1	881	10	US-09-816-860A-2	Sequence 2, Appli
34	115.5	5.1	171	10	US-09-864-761-34988	Sequence 34988, A
35	115.5	5.1	171	10	US-09-864-761-36944	Sequence 36844, A
36	114.5	5.0	248	10	US-09-879-957-40	Sequence 40, Appl
37	114.5	5.0	509	10	US-09-879-957-194	Sequence 194, App
38	114.5	5.0	530	9	US-09-764-868-738	Sequence 738, App
39	114	5.0	550	9	US-09-976-740-47	Sequence 47, Appl
40	114	5.0	550	12	US-10-023-529-47	Sequence 47, Appl
41	114	5.0	550	12	US-10-023-523-47	Sequence 47, Appl
42	113.5	5.0	380	10	US-09-823-240-10	Sequence 10, Appl
43	113	4.9	983	10	US-09-912-917-2	Sequence 2, Appli
44	112.5	4.9	3256	10	US-09-919-172-98	Sequence 98, Appl
45	111.5	4.9	398	10	US-09-896-852-54	Sequence 54, Appl

ALIGNMENTS

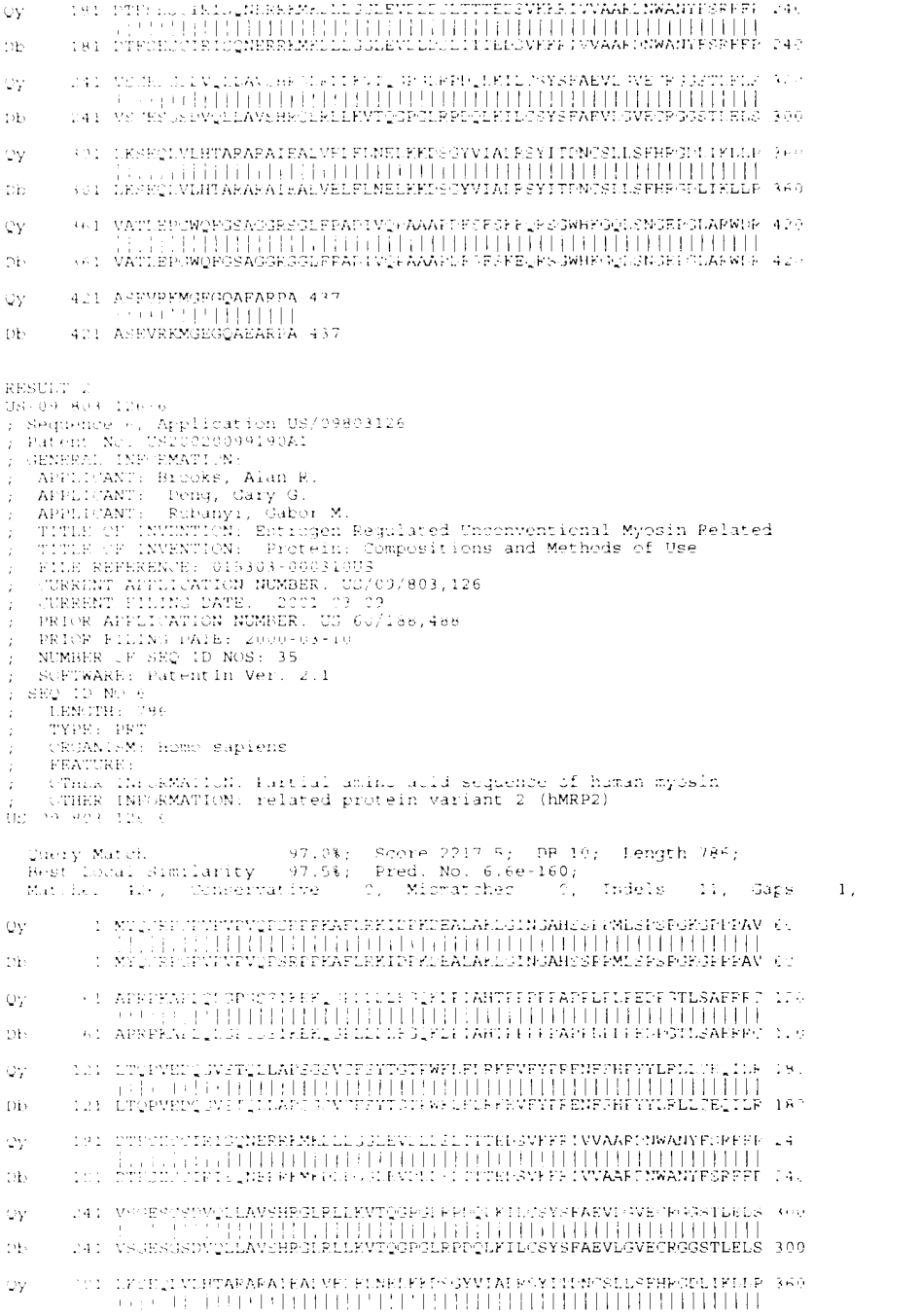
RESULT 1
US-09-803-126-4
; Sequence 4, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Pubanyi, Gabor M
; TITLE OF INVENTION: Estrogen-Regulated Unconventional My sin-Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 1 (hMRP1)
US-09-803-126-4

Query Match 100.0%; Score 2287; DB 10; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.7e-165;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MYQSRPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSFEMLSPSPGKGPPPAV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MYQSRPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSFEMLSPSPGKGPPPAV 60

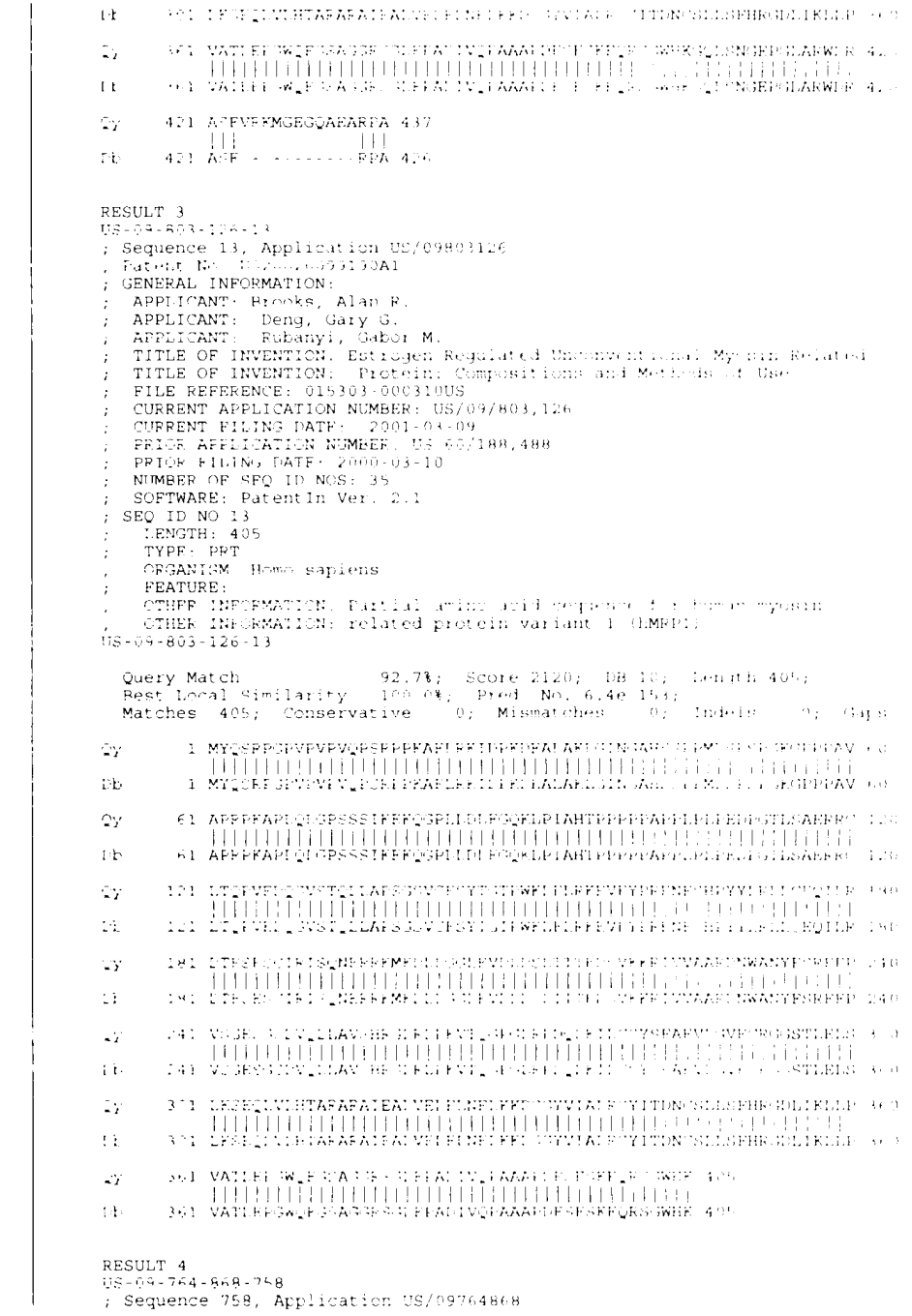
Qy      61 APRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPAPPLPLPEDPGTLSAERRG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 APRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPAPPLPLPEDPGTLSAERRG 120

Qy     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSHPYYLRLLCEQILR 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSHPYYLRLLCEQILR 180
```



RESULT 2
US-09-803-126-6
; Sequence 6, Application US/09803126
; Patent No. US20020009190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 2 (hMRP2)
US-09-803-126-6

Query Match 97.0%; Score 2217.5; DB 10; Length 786;
Best Local Similarity 97.5%; Pred. No. 6.6e-160;
Matches 764; Conservative 0; Mismatches 6; Indels 11; Gaps 1;



RESULT 3
US-09-803-126-13
; Sequence 13, Application US/09803126
; Patent No. US20020009190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 1 (hMRP1)
US-09-803-126-13

Query Match 92.7%; Score 2120; DB 10; Length 405;
Best Local Similarity 100.0%; Pred. No. 6.4e-154;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
US-09-764-868-758
; Sequence 758, Application US/09764868

```
Patent No. US20020168711A1
GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: PTZ32
 CURRENT APPLICATION NUMBER: US/09/764,868
 CURRENT FILING DATE:  2001-01-17
 Prior application data removed - refer to PALM or file wrapper
 NUMBER OF SEQ ID NOS: 1510
 SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 758
;   LENGTH: 333
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-868-758

  Query Match           75.4%;  Score 1725;  DB 9;  Length 333;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-123;
  Matches  333;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     105 LPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPFEN 164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 LPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLPKEVFYPREN 60

Qy     165 FSHPYYLRLLCEQILRDTFSESCIRISQNERRKMKDLLGGLEVDLDSLTTTEDSVKKFIV 224
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 FSHPYYLRLLCEQILRDTFSESCIRISQNERRKMKDLLGGLEVDLDSLTTTEDSVKKFIV 120

Qy     225 VAARDNWANYFSFFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFA 284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAARDNWANYFSRFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFA 180

Qy     285 EVLGVECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDN 344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 EVLGVECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDN 240

Qy     345 CSLLSFHRGDLIKLLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWH 404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CSLLSFHRGDLIKLLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWH 300

Qy     405 KGQLSNGEPGLARWDRASEVRKMGEGQAEARPA 437
           |||||||||||||||||||||||||||||||||
Db     301 KGQLSNGEPGLAPWDRASEVRKMGEGQAEARPA 333


RESULT 5
US-09-803-126-1
; Sequence 1, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
;  APPLICANT: Brooks, Alan R.
;  APPLICANT:  Deng, Gary G.
;  APPLICANT:  Rubanyi, Gabor M.
;  TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
;  TITLE OF INVENTION:  Protein: Compositions and Methods of Use
;  FILE REFERENCE: 015303-000310US
;  CURRENT APPLICATION NUMBER: US/09/803,126
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: US 60/188,488
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1337
;   TYPE: PRT
;   ORGANISM: Mus sp.
;   FEATURE:
;   OTHER INFORMATION: mouse myosin related protein (MPP)
US-09-803-126-1

  Query Match           67.5%;  Score 1543.5;  DB 10;  Length 1337;
  Best Local Similarity  71.3%;  Pred. No. 1.2e-108;
  Matches  308;  Conservative   34;  Mismatches   77;  Indels   13;  Gaps    3;

Qy       1 MYQSRPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60
           |||||||||| |||||:|| | | :| |||||||||||||||| |  |  ||: ||  ||||
Db     914 MYQSRPGPVAVPVQPTRPIKTFQKKNDPKDEALAKLGINGVH-LPLSTSPNQGKSSPPAV 972

Qy      61 APPPKAPLQLGPSSSIKEKQGPLLDLFG---QKLPIAHTPPPPPAPPLPLPEDPGTLSAE 117
           |||||  .| || ||.|||||| |||        | |  |||||| | ||  :| | | |
Db     973 VFRPKAPPRLEPSLSIQEKQGPLRDLFGPCSPNPPTAPAPPPPPALFFFLGGEPPTPSVE 1032

Qy     118 RRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSHPYYLRLLCEQ 177
           || |.||: .||:|| ||||||||  |||||||||||||||||||| | |||:|
Db    1033 SHALTEPMEDKNISTKLLVPSGSVCFSYAGAPWKLFLPKEVFYPRENFSHPYCLSLLCQQ 1092

Qy     178 ILRDTFSESCIRISQNERRKMKDLLGGLEVDLDSLTTTEDSVKKFIVVAARDNWANYFSR 237
           |||||.||| ||||:|| ||| ||| |..|   |||:||||||||||||||||||||||
Db    1093 ILRDTFTESCTRISQDERHKMKGLLGDLEVSLETLDIVEDSIKKPIVVAARDNWANYFSR 1152

Qy     238 FFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECRGGSTL 297
           |||||||||||||| ||||||||||| |    |||| ||||:|||| |:||| |||
Db    1153 IFPVSGESGSDVQLLGVSHRGLRLLKVTQSPSFHLDQLKTLCSYSYAEVLTVQCRGRSTL 1212

Qy     298 ELSLKSEQLVLHTAPAPAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIK 357
           |||||:|||:|||| |||:|:: |||:|:|||||||||||||||: ||||||||||
Db    1213 ELSLKNEQLILHTAWAPAIKAMVDLFLSELRKDSGYVIALRSYITDLNSLLSFHPGDLIR 1272

Qy     358 LLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAR 417
           |||| |||||||||||||||| |:|||||| |||  :|: | |  :              ::
Db    1273 LLPVTALEPGWQFGSAGPSGLFPDDVVQPAAAPDLSFSLGKRNSWQK---          PSK 1323

Qy     418 WDRASEVRKMGE 429
           | |||| |
Db    1324 LGPAQEVRKTEE 1335

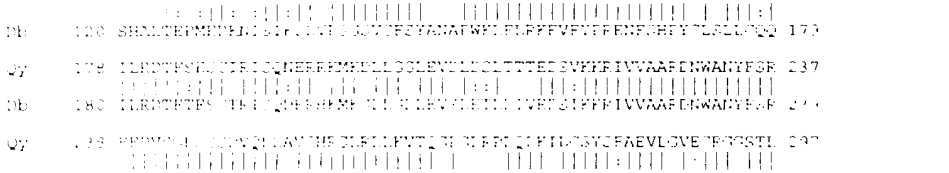

RESULT 6
US-09-803-126-12
; Sequence 12, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
;  APPLICANT: Brooks, Alan R.
;  APPLICANT:  Deng, Gary G.
;  APPLICANT:  Rubanyi, Gabor M
;  TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
;  TITLE OF INVENTION:  Protein: Compositions and Methods of Use
;  FILE REFERENCE: 015303-000310US
;  CURRENT APPLICATION NUMBER: US/09/803,126
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: US 60/188,488
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 407
;   TYPE: PRT
;   ORGANISM: Mus sp.
;   FEATURE:
;   OTHER INFORMATION: Partial amino acid sequence for mouse myosin
;   OTHER INFORMATION. related protein (mMRP)
US-09-803-126-12

  Query Match           67.2%;  Score 1537;  DB 10;  Length 407;
  Best Local Similarity  74.0%;  Pred. No. 7.5e-109;
  Matches  302;  Conservative   32;  Mismatches   70;  Indels    4;  Gaps    2;

Qy       1 MYQSRPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60
           |||||||||| |||||:|| | | :| |||||||||||||||| |  |  ||: ||  ||||
Db       1 MYQSRPGPVAVPVQPTRPIKTFQKKNDPKDEALAKLGINGVH LPLSTSPNQGKSSPPAV 59

Qy      61 APRPKAPLQLGPSSSIKEKQGPLLDLFG---QKLPIAHTPPPPPAPPLPLPEDPGTLSAE 117
           |||||  .| || ||.|||||| ||||        | |  |||||| | ||  :| | | |
Db      60 VFRPKAPPRLEPSLSIQEKQGPLRDLFGPCSPNPPTAPAPPPPPALPPPLSGEPKTPSVE 119

Qy     118 RRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPRENFSHPYYLRLLCEQ 177
```



```
RESULT 7
US-09-864-761-40685
; Sequence 40685, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT: Funk, David E.
;  APPLICANT: Hanzel, David K.
;  APPLICANT: Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE: 2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,450
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 60/232,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 60/214,478
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,283
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Aeomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40685
;  LENGTH: 56
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AC019214.2
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;  OTHER INFORMATION: SWISSPROT HIT: Q99663, EVALUE = 5.5e-10
;  OTHER INFORMATION: EST_HUMAN HIT: AW249878.1, EVALUE = 1.7e-28
US-09-864-761-40685

  Query Match             14.4%;  Score 306;  DB 10;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-17;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    368 WQFGSAGGRSGLFPADIVGPAAAPDPSPSKEQPGGWHK 405
          ||||||||||||||||||||||||||||||||||||||||
Db      1 WQFGSAGGRSGLFPADIVGPAAAPDPSPSKEQPGGWHK 38


RESULT 8
US-09-858-155A-2
; Sequence 2, Application US/09858155A
; Patent No. US20020137049A1
; GENERAL INFORMATION:
;  APPLICANT: Mark, Robert
;  APPLICANT:  Young, Kathleen H.
;  APPLICANT:  Wood, Andrew
;  TITLE OF INVENTION: PARG9, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND
;  TITLE OF INVENTION:  USES RELATED THERETO
;  FILE REFERENCE: GNN-005
;  CURRENT APPLICATION NUMBER: US/09/858,155A
;  CURRENT FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;  LENGTH: 559
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-858-155A-2

  Query Match              6.3%;  Score 143;  DB 10;  Length 559;
  Best Local Similarity   24.5%;  Pred. No. 0.0026;
  Matches   67;  Conservative   36;  Mismatches   97;  Indels   74;  Gaps   15;

Qy      1 MTLQEI      GIVIVIVLVLFSIFFAFLRKILIFPDIALAPLEIINANDSS- PPMLSPSPG 54
          : ::||     |  || | |: ||        |   ||:   :  : :|  _| : : |
Db    313 LIENRIIPPATDPTFVPVCFTPPFFP     PFLPWALSTSSLPAVMTCTTPPPTVPFPT  369

Qy     54 PTPPPAVA--    -PPPAPLQIGP           PPPIPPPQVPCIIPPPQKLPIAHT  96
            |||| |       | | ||||: |          : :  :   :      :  :  :
Db    357   PFPATALQAPAVFFPPAPLQIAPGVLHPAPPPIAPPLVQPSPPV    APAAPVVTTV 410

Qy     97 --------      PPPPPAPPLP    LPEDPGTLSAEPPPCLTQPVEVQVV PTQLLAPG 150
                       ||||| ||||     |  | |:|    :  :   :  :  |  ||
Db    411 FVHFIFQGEVQGIPFFFFFFIFFIGIFIGIVIVIA    IAIP  FGGRIFISIAPG 464

Qy    151 SEVIFSTTSTIWPLIFPFVVFYFPEDFLHFYYLRLLCE -QILRDTFSEPTTRINQNERR 209
            |       |       :|    |  ||  | : :   :|  :   :          :||
Db    465 FHVFLMFFSTT       SQVIFAGEPESHPGIILPVIICAFVLLEATEFIQLIKPVEEQR 518

Qy    197 PMPILGSNIEVFIGSLITIPSVPFFIVVAAPVN 230
          :.    :| |: ..      : :|| |   |:
Db    519 EQEAPHEFIENFVATI------IGPPIAVIVSDS 546


RESULT 9
US-10-001-873-50
```

Sequence 50, Application US/10001873
Patent No. US20020160388A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
APPLICANT: Turner, Leah
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
FILE REFERENCE: DEX-0275
CURRENT APPLICATION NUMBER: US/10/001,873
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/252,055
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: 60/252,496
PRIOR FILING DATE: 2000-11-20
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn version 3.1
SEQ ID NO 50
LENGTH: 1134
TYPE: PRT
ORGANISM: Homo sapien
US-10-001-873-50

Query Match 5.7%; Score 131.5; DB 9; Length 1134;
Best Local Similarity 34.3%; Pred. No. 0.049;
Matches 48; Conservative 4; Mismatches 65; Indels 23; Gaps 6;

Qy 4 SRPGPVPVPVQPS----------RPPKAFLRKIDPKDEALAKLGIN---GAHSSPPML 48
Db 886 SAAAPAPVSIAPATKTSPLAPGGPLSPRKDAGPAKDPKDPPKPVKKGKEAGPGAGLPPPRA 945

Qy 49 S--PSPGKGPPPAVAPRPKAPLQLGPSSSIKEKQGPLL--DLFGQKLPIAHTPPPPPAPP 104
Db 946 PALPSEARAPPPPPPPPPHPPL---PPPPLPPPPLPLRLPPLPPPPLPRPHPPPPPPLPP 1002

Qy 105 LPLPEDPGTLSAERRCLTQP 124
Db 1003 LLPPPQTRTLPA ARTMPQP 1021

RESULT 10
US-09-853-386-69
; Sequence 69, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Evelyn
; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR
; TITLE OF INVENTION: Subfamily of Nuclear Transcription Factors
; FILE REFERENCE: P01972US1
; CURRENT APPLICATION NUMBER: US/09/853,386
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Sus scrofa
US-09-853-386-69

Query Match 5.6%; Score 129; DB 10; Length 446;
Best Local Similarity 28.5%; Pred. No. 0.022;
Matches 51; Conservative 14; Mismatches 62; Indels 52; Gaps 8;

Qy 3 QSRPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPG---KGPPPA 59
Db 115 QQQQQPPPQQQQPSIPPPS-----GPEDEVL---------------PSTSMYFKQSPPS 153

Qy 60 VAPRPKAPLQLG-------PSSSIKEKQGPLLD--------LFGQKLPIAHTPPPPPAPP 104
Db 154 TPTTPVFPQQAGALWEDALPSAQGCIAPGPLLDPPMKAVPTVAGARFPLPHFKPSPPHPP 213

Qy 105 LPLPE-------DPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRK 156
Db 214 APSPAGGHHLGYDPTAAAA----LGLPL--GAAAAAAAAAAGSQAAALEGHFYGLPLAK 266

RESULT 11
US-09-853-386-70
; Sequence 70, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Evelyn
; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR
; TITLE OF INVENTION: Subfamily of Nuclear Transcription Factors
; FILE REFERENCE: P01972US1
; CURRENT APPLICATION NUMBER: US/09/853,386
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Sus scrofa
US-09-853-386-70

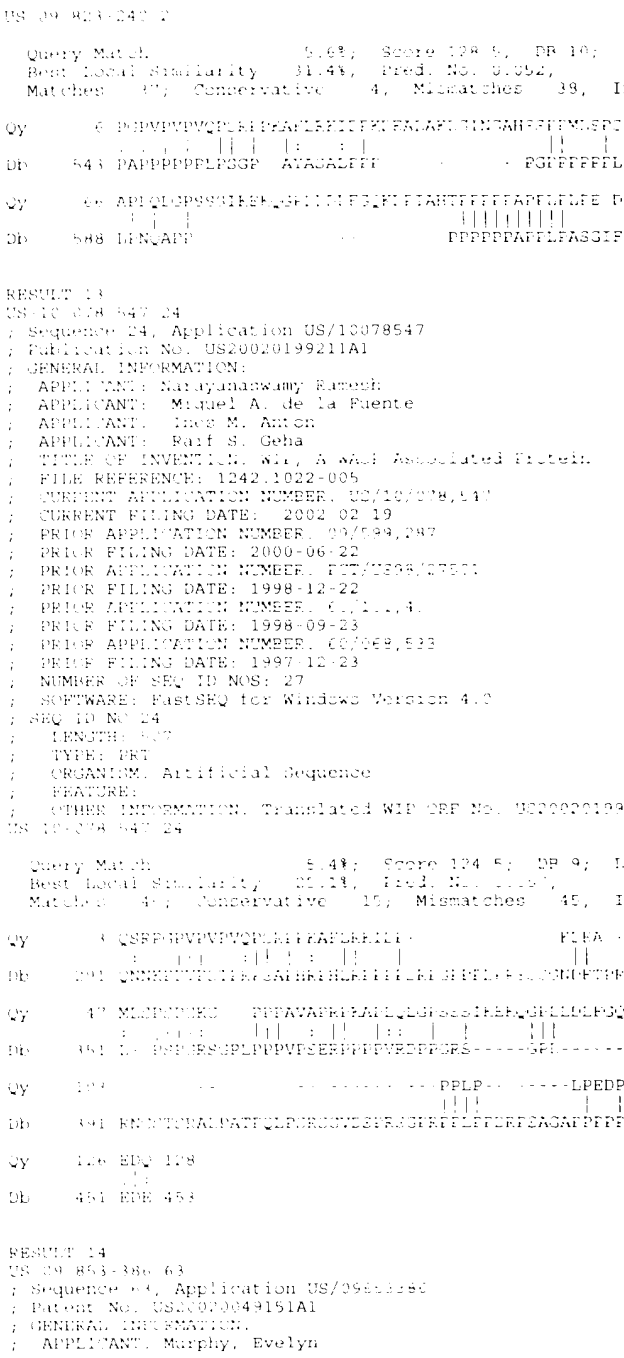
Query Match 5.6%; Score 129; DB 10; Length 643;
Best Local Similarity 28.5%; Pred. No. 0.035;
Matches 51; Conservative 14; Mismatches 62; Indels 52; Gaps 8;

Qy 3 QSRPGPVPVPVQPSRPPKAFLRKIDPKDEALAKLGINGAHSSPPMLSPSPG---KGPPPA 59
Db 115 QQQQQPPPQQQQPSIPPPS-----GPEDEVL---------------PSTSMYFKQSPPS 153

Qy 60 VAPRPKAPLQLG-------PSSSIKEKQGPLLD--------LFGQKLPIAHTPPPPPAPP 104
Db 154 TPTTPVFPQQAGALWEDALPSAQGCIAPGPLLDPPMKAVPTVAGARFPLPHFKPSPPHPP 213

Qy 105 LPLPE-------DPGTLSAERRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRK 156
Db 214 APSPAGGHHLGYDPTAAAA----LGLPL--GAAAAAAAAAAGSQAAALEGHFYGLPLAK 266

RESULT 12
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E.Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureio
; TITLE OF INVENTION. Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus



; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
; TITLE OF INVENTION: Therapeutic Approaches to Disorders by Suppression of the NFkB
; TITLE OF INVENTION: Subfamily of Nuclear Transcription Factors
; FILE REFERENCE: PC1972US1
; CURRENT APPLICATION NUMBER: US/09/853,386
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63
; LENGTH: 625
; TYPE: PRT
; ORGANISM: HUMAN
US-09-853-386-63

Query Match 5.4%; Score 124.5; DB 10; Length 625;
Best Local Similarity 30.2%; Pred. No. 0.075;
Matches 49; Conservative 16; Mismatches 92; Indels 45; Gaps 7;

Qy 14 QPSPPPFAFLPPIDPPIPALAPIGINGAHSSPPMLSPSPSRGHPAVA 63
Db 114 QPSIPPAS-----SPEDEVLPSIDMYFPQSPPS TPTTPAPPPLAVALWLPEALPSAPST 160

Qy 70 LGPSSSIKEPLGPLLD--------LPGQKLPIAHTPPPPPAPPLPLPE 109
Db 161 IAP--------GPLLDPPMKAVPTVAGAPFPIPHPPPSPPHPAPSPAPSMPHLGYPTAA 210

Qy 110 SAPPPLTLPVEPLNVSTPLLAPSGVPFSYTSTHWPLPLKR 150
Db 211 AA PIPI PAA PAAPGPAAAPPHPPPPIPAP 249

RESULT 15
US-09-853-386-64
; Sequence 64, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Evelyn
; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
; TITLE OF INVENTION: Therapeutic Approaches to Disorders by Suppression of the NFkB
; TITLE OF INVENTION: Subfamily of Nuclear Transcription Factors
; FILE REFERENCE: PC1972US1
; CURRENT APPLICATION NUMBER: US/09/853,386
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 625
; TYPE: PRT
; ORGANISM: HUMAN
US-09-853-386-64

Query Match 5.4%; Score 124.5; DB 10; Length 625;
Best Local Similarity 30.2%; Pred. No. 0.075;
Matches 49; Conservative 16; Mismatches 92; Indels 45; Gaps 7;

Qy 14 QPSPPPFAFLPPIDPPIPALAPIGINGAHSSPPMLSPSPSRGHPAVA 63
Db 114 QPSIPPAS-----SPEDEVLPSIDMYFPQSPPS TPTTPAPPPLAVALWLPEALPSAPST 160

Qy 70 LGPSSSIKEPLGPLLD--------LPGQKLPIAHTPPPPPAPPLPLPE 109
Db 161 IAP--------GPLLDPPMKAVPTVAGAPFPIPHPPPSPPHPAPSPAPSMPHLGYPTAA 210

Qy 110 SAPPPLTLPVEPLNVSTPLLAPSGVPFSYTSTHWPLPLKR 150
Db 211 AA PIPI PAA PAAPGPAAAPPHPPPPIPAP 249

b 219 AA----LSLPL---GAA----AAAGSQAAALEGHPYGLPLAK 249

Search completed: January 10, 2003, 02.29:06
Job time : 36.7992 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: January 9, 2003, 03:03:27 ; Search time 10149.4 Seconds
(without alignments)
11968.660 Million cell updates/sec

Title: US-09-803-126-5
Perfect score: 4174
Sequence: 1 cggcagcagcaggctcgggc ... .aaaaaaaaaaaaaaaaaaa 4174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39 em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

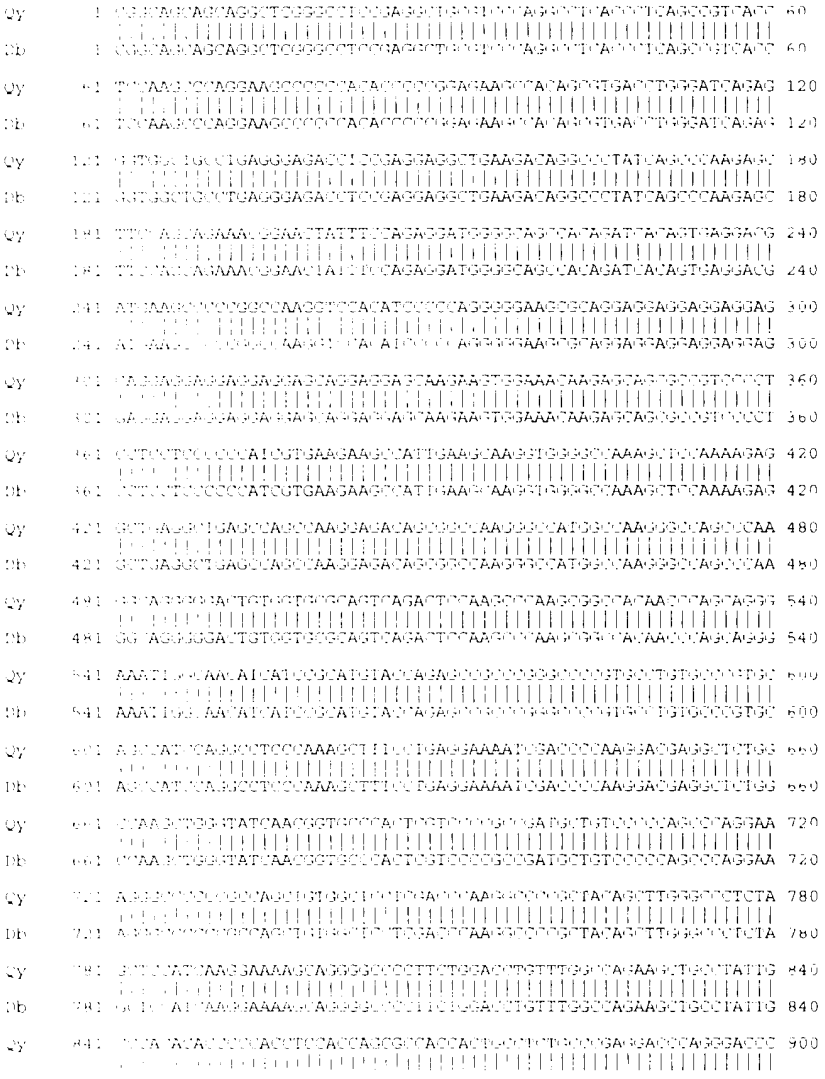
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4174	100.0	4174	6	AX253082	AX253082 Sequence
2	3562	85.3	4602	9	AF418292	AF418292 Homo sapi
3	3352	80.3	3780	6	AX253084	AX253084 Sequence
4	2764	66.2	4208	9	AF418293	AF418293 Homo sapi
5	2246	53.8	2347	9	BC027875	BC027875 Homo sapi
6	2102.4	50.4	6004	9	AB058686	AB058686 Homo sapi
7	2088.6	50.0	2372	9	AK097896	AK097896 Homo sapi
8	1855.4	44.5	1891	9	AK026339	AK026339 Homo sapi
9	1787.6	42.8	6293	6	AX253079	AX253079 Sequence
10	1751.2	42.0	1767	9	HSM803107	AL713730 Homo sapi
11	1691.4	40.5	7925	9	AF418290	AF418290 Homo sapi
12	1386.6	33.2	1998	9	AK096869	AK096869 Homo sapi
13	1376.8	33.0	2099	9	AK025259	AK025259 Homo sapi
14	1135.6	27.2	7515	9	AF418286	AF418286 Homo sapi
15	946.4	22.7	4375	6	AX253080	AX253080 Sequence
16	775.6	18.6	1547	10	BC024506	BC024506 Mus muscu
c 17	715	17.1	215424	2	AC019214	AC019214 Homo sapi
18	715	17.1	217584	2	AC087749	AC087749 Homo sapi
c 19	708.6	17.0	170056	2	AC087643	AC087643 Homo sapi
20	542.2	13.0	178515	2	AC015551	AC015551 Homo sapi
21	442	10.6	170056	2	AC087643	AC087643 Homo sapi
22	293	7.0	5709	9	AF418289	AF418289 Homo sapi
c 23	292	7.0	311	6	AX330132	AX330132 Sequence
c 24	292	7.0	311	6	AX332694	AX332694 Sequence
c 25	292	7.0	311	6	AX333379	AX333379 Sequence
c 26	292	7.0	311	6	AX408873	AX408873 Sequence
27	214.6	5.1	11876	9	AF144094	AF144094 Homo sapi
28	201.4	4.8	5351	10	AF053130	AF053130 Mus muscu
29	201.4	4.8	11769	10	AF144095	AF144095 Mus muscu
30	134.2	3.2	147804	10	AL645647	AL645647 Mouse DNA
c 31	134.2	3.2	260310	2	AL645852	AL645852 Mus muscu
c 32	125.2	3.0	180956	2	AC123410	AC123410 Rattus no
33	125.2	3.0	200832	2	AC130970	AC130970 Rattus no
34	109	2.6	3110	9	HSM803101	AL713794 Homo sapi
c 35	73.4	1.8	7218	6	I66494	I66494 Sequence 14
c 36	73.4	1.8	226714	2	AC079532	AC079532 Mus muscu
37	72.4	1.7	125020	9	AF429315	AF429315 Homo sapi
38	71.8	1.7	202370	2	AC102899	AC102899 Mus muscu
39	70.6	1.7	186400	2	AC095843	AC095843 Rattus no
40	69.2	1.7	207683	2	AC098712	AC098712 Mus muscu
c 41	69	1.7	123380	2	AC126533	AC126533 Rattus no
42	68.8	1.6	139838	2	AC128371	AC128371 Rattus no
43	68.4	1.6	126374	9	HS73J07	AL353993 Homo sapi
44	67.8	1.6	69070	2	AC099959	AC099959 Mus muscu
45	67.8	1.6	169827	2	AC111122	AC111122 Mus muscu

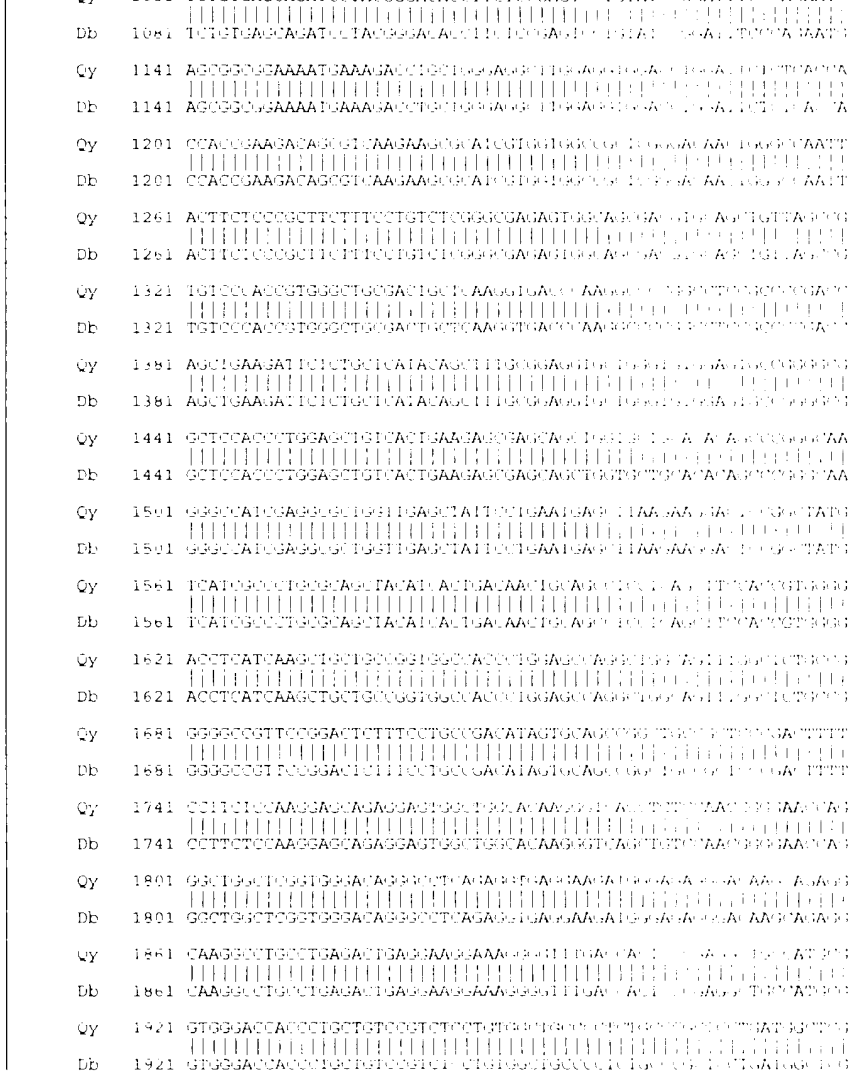
ALIGNMENTS

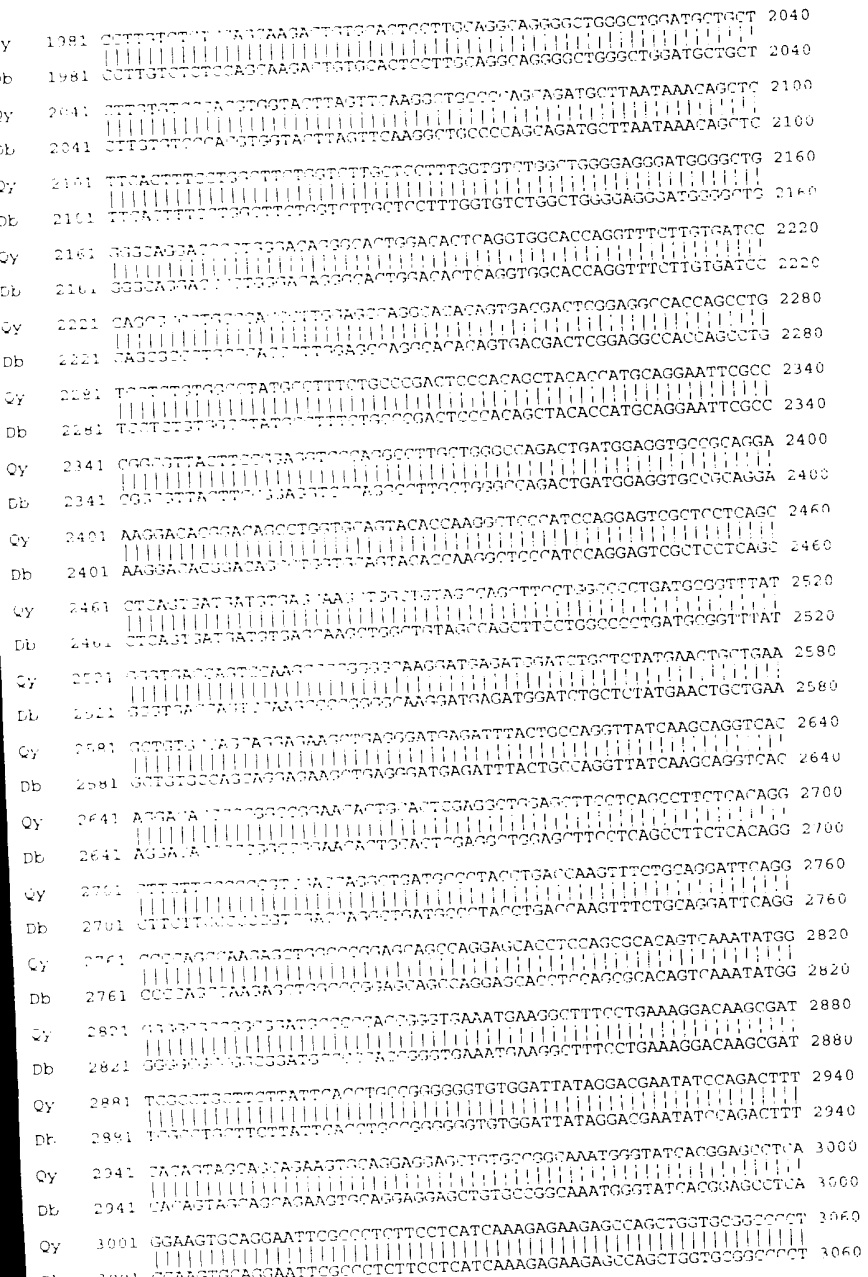
RESULT 1
AX253082
LOCUS AX253082 4174 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168866.
ACCESSION AX253082
VERSION AX253082.1 GI:15986279
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4174)
AUTHORS Brooks,A.P., Deng,G.G. and Rubanyi,G.M.
TITLE Estrogen-regulated unconventional myosin related protein: compositions and methods of use

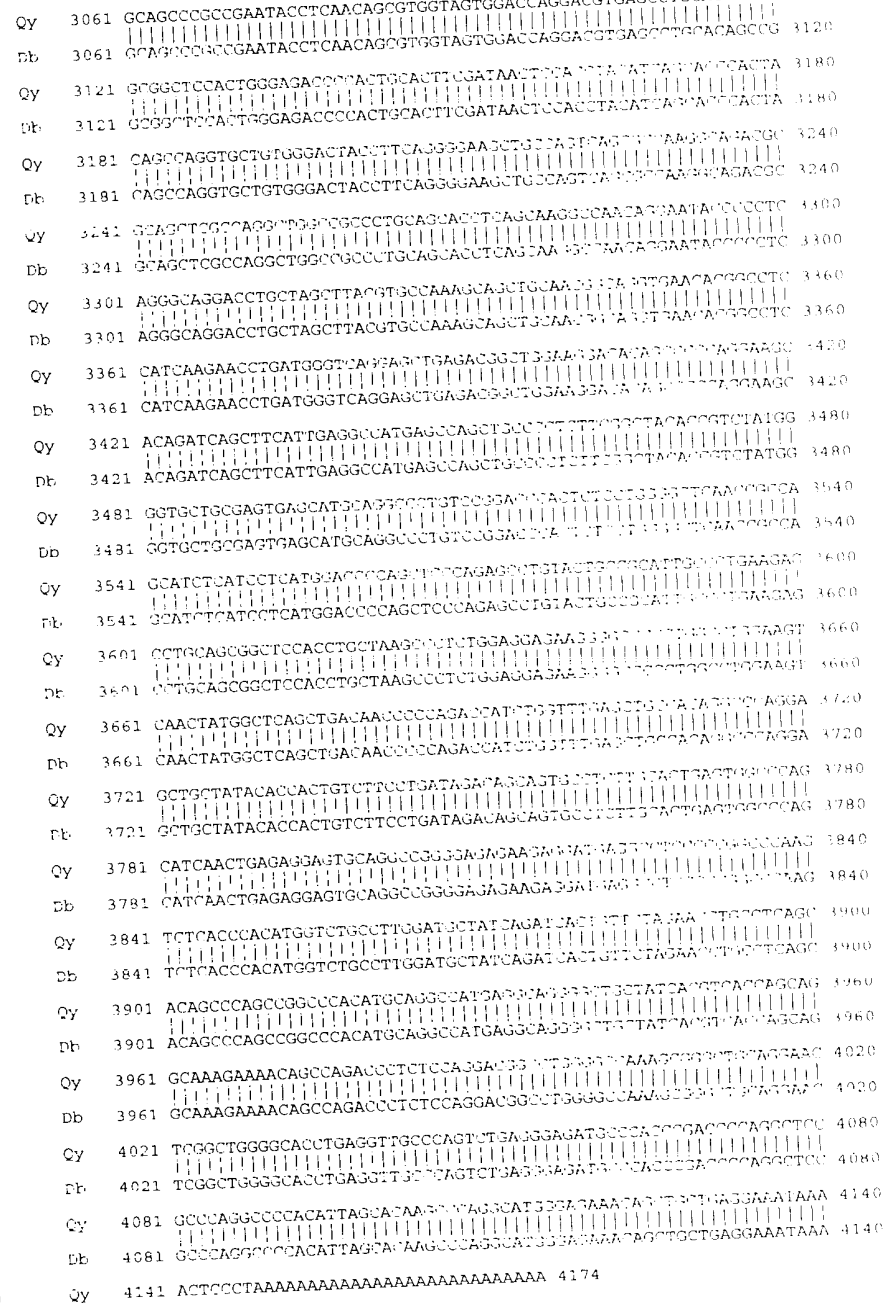
JOURNAL Patent: WO 0168866-A 5 20-SEP-2001;
SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1..4174
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Partial DNA sequence for human myosin related protein variant 1 (hMRP1)"
BASE COUNT 901 a 1339 c 1237 g 697 t
ORIGIN

Query Match 100.0%; Score 4174; DB 6; Length 4174;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;









```
Db   4141 ACTTCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4174


RESULT 2
AF418292
LOCUS       AF418292                1602 bp    mRNA    linear   PRI 06-JUN-2002
DEFINITION  Homo sapiens clone cDNA1 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418292
VERSION     AF418292.1  GI:16306915
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4602)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21960014
    PUBMED  11964073
REFERENCE   2  (bases 1 to 4602)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13 SEP 2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..4602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="cDNA1"
                     /tissue_type="kidney; stomach"
     misc_feature    <1..>4602
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT      927 a   1495 c   1405 g    775 t
ORIGIN

  Query Match             85.3%;  Score 3562;  DB 9;  Length 4602;
  Best Local Similarity   99.8%;  Pred. No. 0,
  Matches 3508;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    3;

Qy       1 CCCCAGCACAGAGGCTCGGGGCCTCCGAGGCTGCCCCCCAGGCCTCACCCTCACCCCGTCACC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     999 CGGTAGCAGCAGGCTCGGGGCTCCGAGGCTGCCTCCCAGGCCTCACCCTCACCCCGTCACC 1058

Qy      61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1059 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 1118

Qy     121 GGTGGATGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1119 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 1178

Qy     181 TTTCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1179 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 1238

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1239 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG 1298

Qy     301 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGTGGAAACAAGAGCAGTGCCGTCCCCT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1299 GAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGTGGAAACAAGAGCAGTGCCGTCCCCT 1358

Qy     361 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1359 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 1418

Qy     421 GCTGAGGCTGAGCCAGCCAAGGAGACAGCCGCCAAGGGCATCCCCAAGGCTCAGCCAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1419 GCTGAGGCTGAGCCAGCCAAGGAGACAGCCGCCAAGGGCATCCCCAAGGCTCAGCCAA 1478

Qy     481 CCCACCCCCGACCTGGCCCCCACTCACACCTCCAAGCCCAAGGCCCACAACCTACCAGTAGG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1479 GGCAGGGGGACCTGTGGTGGCGCAGCTCAGCTCCAAGCCCAAGGCCCACAACCTACCAGTAGG 1538

Qy     541 GAAAATTGGCAACATCATCCGCATGTACCAGAGCCTGCCCCCCGGCGTGCCCGTGTCCTG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1539 GAAAATTGGCAACATCATCCGCATGTACCAGAGCCTGCCCCCCGGCGTGCCCGTGTCCTG 1598

Qy     601 CAGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGAGCCCAAAGAGGAGGTCTCG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1599 CAGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGAGCCCAAAGAGGAGGTCTCG 1658

Qy     661 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGTGAGCCCCCAGCCCAGGA 715
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1659 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGTGAGCCCCCAGCCCAGGA 1713

Qy     716 AAGGGCCCCCCGCCAGCTGTGGCTCCTCGACCCAAGGCCCCCACAGCTCCCCTCT 775
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1714 AAGGGCCCCCCGCCAGCTGTGGCTCCTCGACCCAAGGCCCCCACAGCTCCCCTCT 1773

Qy     776 ACCCTCACCAAAGAAAACAAGGGGCCCCTTCTGGACCCTGTTTCCAGAAGCTCCCATT 835
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1774 AGCTCCATCAAGGAAAACAAGGGGCCCCTTCTGGACCCTGTTTCCAGAAGCTCCCATT 1833

Qy     836 GCCCACACTACCCCCACCTCCACCAGCGCCACCACTGCCTCTCCCCCAGAGACCCAGGACCT 895
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1834 GCCCACACTACCCCCACCTCCACCAGCGCCACCACTGCCTCTCCCCCAGAGACCCAGGACCT 1893

Qy     896 CTTTCAGCAGAGGGTCGTTGCTTGACACAGCCCGTGGAGGACCAAGGGGCTCCACCCAGT 955
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1894 CTTTCAGCAGAGGGTCGTTGCTTGACACAGCCCGTGGAGGACCAAGGGGCTCCACCCAGT 1953

Qy     956 CTACTGGGGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGCAGCCCTGGAAGTTCTTC 1015
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1954 CTACTGGGGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGCAGCCCTGGAAGTTCTTC 2013

Qy     1016 CTACGCAAGGAGGTGTTCTACCCACGGGAGAAGTTCAAGATCCTGAGGTTC 1075
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2014 CTACGCAAGGAGGTGTTCTACCCACGGGAGAAGTTCAAGATCCTGAGGTTC 2073

Qy     1076 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGCAAGAATCCTAAAT 1135
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2074 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGCAAGAATCCTAAAT 2133

Qy     1136 GAGGGGCGGAAAATGAAAGACCTGCTGGAGGCTTGGAGGCTGGACTTCCAGATCTCCAGT 1195
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2134 GAGGGGCGGAAAATGAAAGACCTGCTGGAGGCTTGGAGGCTGGACTTCCAGATCTCCAGT 2193

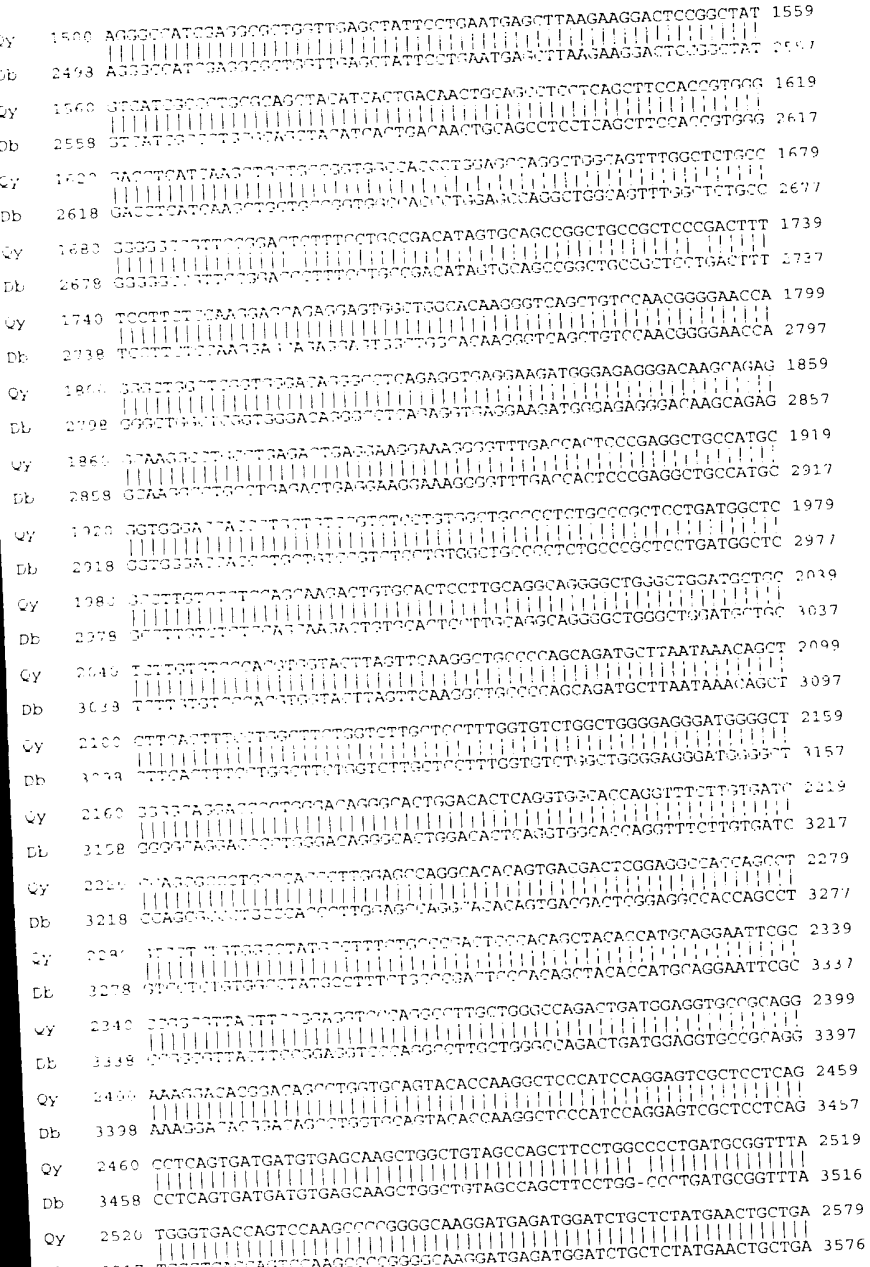
Qy     1196 ACCACCGAAGACAGCGTCAAGAAGCGCATCGTGGTGGCCGCTTGGACAAACTGGGTAAT 1255
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2194 ACCACCGAAGACAGCGTCAAGAAGCGCATCGTGGTGGCCGCTTGGACAAACTGGGTAAT 2253

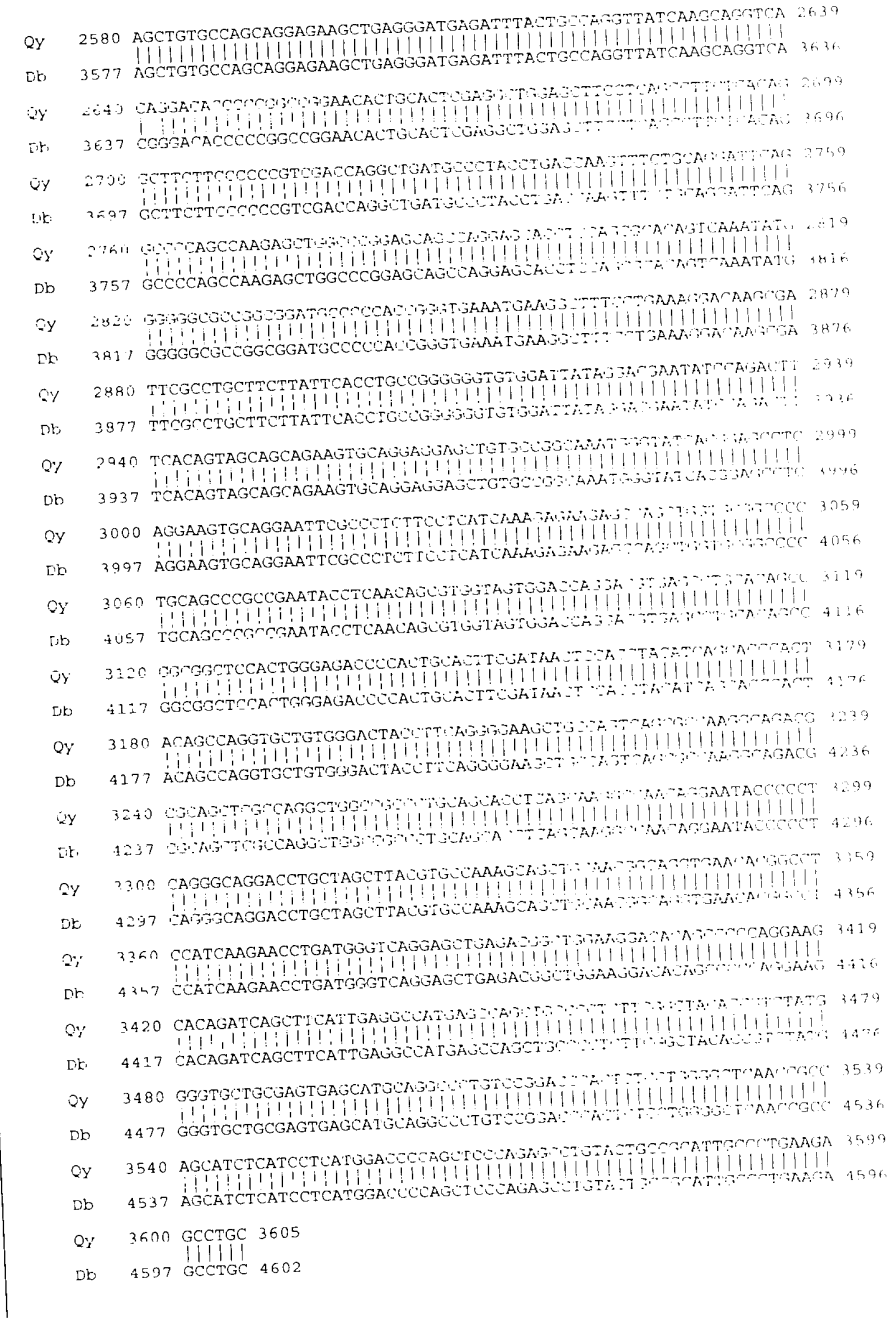
Qy     1256 TACTTCTCCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGGAACCCCAACTGTAGCT 1315
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2254 TACTTCTCCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGGAACCCCAACTGTAGCT 2313

Qy     1316 GTGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGGCCCCCCCGAC 1365
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2314 GTGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGGCCCCCCCGAC 2363

Qy     1366 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCCGAGGTGCTGGGTTTTGACTTTGGGGC 1425
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2364 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCCGAGGTGCTGGGTTTTGACTTTGGGGC 2423

Qy     1426 GGCTCCACCCTGGAGCTGTCACTGAAGAGCGGAGCAGCTGGTGCTCTACACAGCCCTGA 1485
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2424 GGCTCCACCCTGGAGCTGTCACTGAAGAGCGGAGCAGCTGGTGCTCTACACAGCCCTGA 2483
```





```
RESULT 3
AX253084
LOCUS       AX253084        3780 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 7 from Patent WO0168866.
ACCESSION   AX253084
VERSION     AX253084.1  GI:15986280
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3780)
  AUTHORS   Brooks,A.R., Deng,G.G. and Rubanyi,G.M.
  TITLE     Estrogen regulated unconventional myosin related protein
            compositions and methods of use
  JOURNAL   Patent: WO 0168866-A 7 20-SEP-2001;
            SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..3780
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Partial DNA sequence for human myosin related
                     protein variant 2 (hMRP2)"
BASE COUNT      838 a   1228 c   1107 g    607 t
ORIGIN

  Query Match             80.3%;  Score 3352;  DB 6;  Length 3780;
  Best Local Similarity   90.5%;  Pred. No. 0;
  Matches 3778;  Conservative    0;  Mismatches    0;  Indels  396;  Gaps    3;

Qy          1 [illegible]CAGGCTCGGGCTCCGAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACC 60
Db          1 [illegible]CAGGCTCGGGCTCCGAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACC 60

Qy         61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGGTGACCTGGGATCAGAG 120
Db         61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGGTGACCTGGGATCAGAG 120

Qy        121 [illegible]CTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
Db        121 [illegible]CTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180

Qy        181 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
Db        181 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240

Qy        241 ATGAA[illegible]CCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAG 300
Db        241 ATGAA[illegible]CCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAG 300

Qy        301 [illegible]GAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360
Db        301 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360

Qy        361 [illegible]CCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
Db        361 [illegible]CCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420

Qy        421 [illegible]GAGCCAGCCAAGGAGACAGGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480
Db        421 GCTGAGGGCTGAGCCAGCCAAGGAGACAGGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480

Qy        481 GG[illegible]GGACTGTGGTGGCGCAGTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGGG 540
Db        481 GG[illegible]GGACTGTGGTGGCGCAGTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGGG 540

Qy        541 AAATT[illegible]CATGTACCAGAGCCG[illegible]GGGGCCCCGTGCCTGTGCCCCGTGC 600
Db        541 AAATTGGCAACATCATCCGCATGTACCAGAGCCG[illegible]GGGGCCCCGTGCCTGTGCCCCGTGC 600

Qy        601 AG[illegible]CAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGG 660
Db        601 AGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGG 660

Qy        661 CCAAGCTGGGTATCAACGGTGCCCACTCGTCCCC[illegible] 720
Db        661 CCAAGCTGGGTATCAACGGTGCCCACTCGTCCCC[illegible] 720

Qy        721 AGGGCCCCCGCCAGCTGTGGGTCCTCGACCCAAGG[illegible] 780
Db        721 AGGGCCCCCGCCAGCTGTGGGTCCTCGACCCAAGG[illegible] 780

Qy        781 GCTCCATCAAGGAAAAGCAGGGG[illegible]TGCTATT[illegible] 840
Db        781 GCTCCATCAAGGAAAAGCAGGGGCCCCTCTGGACC[illegible]AGAAGTGCTATT[illegible] 840

Qy        841 CCCACACACCCCCACCTCCACCAGGCCCACCACTGCC[illegible] 900
Db        841 CCCACACACCCCCACCTCCACCAGGCCCACCACTGCC[illegible] 900

Qy        901 TTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGGAGGA[illegible] 960
Db        901 TTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGGAGGA[illegible] 960

Qy        961 TACTCGGGCCCTCTGGCAGCGTGTGCTTCTCCTACAC[illegible] 1020
Db        961 TACTCGGGCCCTCTGGCAGCGTGTGCTTCTCCTACAC[illegible] 1020

Qy       1021 TACGCAAGGAGGTGTTCTACCCACGGGAGAACTTCAGCCAT[illegible]TGAGGCTC[illegible] 1080
Db       1021 TACGCAAGGAGGTGTTCTACCCACGGGAGAACTTCAGCCAT[illegible] 1080

Qy       1081 TCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTA[illegible]CAGAATG 1140
Db       1081 TCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTA[illegible]GATTTCCAGAATG 1140

Qy       1141 AGCGGCGGAAAATGAAAGACCTGCTGGGAGGCTTGGAG[illegible] 1200
Db       1141 AGCGGCGGAAAATGAAAGACCTGCTGGGAGGCTTGGAG[illegible] 1200

Qy       1201 CCACCGAAGACAGCGTCAAGAAGCGCATCGTGGTGGCCGCTC[illegible]AACTGGG[illegible]AATT 1260
Db       1201 CCACCGAAGACAGCGTCAAGAAGCGCATCGTGGTGGCCGCTC[illegible]AACT[illegible]AATT 1260

Qy       1261 ACTTCTCCCGCTTCTTCCTGTCTCGGGCGAGAGTGGCAGCGA[illegible]TGTTAGCCG 1320
Db       1261 ACTTCTCCCGCTTCTTCCTGTCTCGGGCGAGAGTGGCAGCGA[illegible]TGTTAGCCG 1320

Qy       1321 TGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGG[illegible] 1380
Db       1321 TGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGG[illegible] 1380

Qy       1381 AGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTG[illegible] 1440
Db       1381 AGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTG[illegible] 1440

Qy       1441 GCTCCACCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTG[illegible]AA 1500
Db       1441 GCTCCACCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGC[illegible]AA 1500

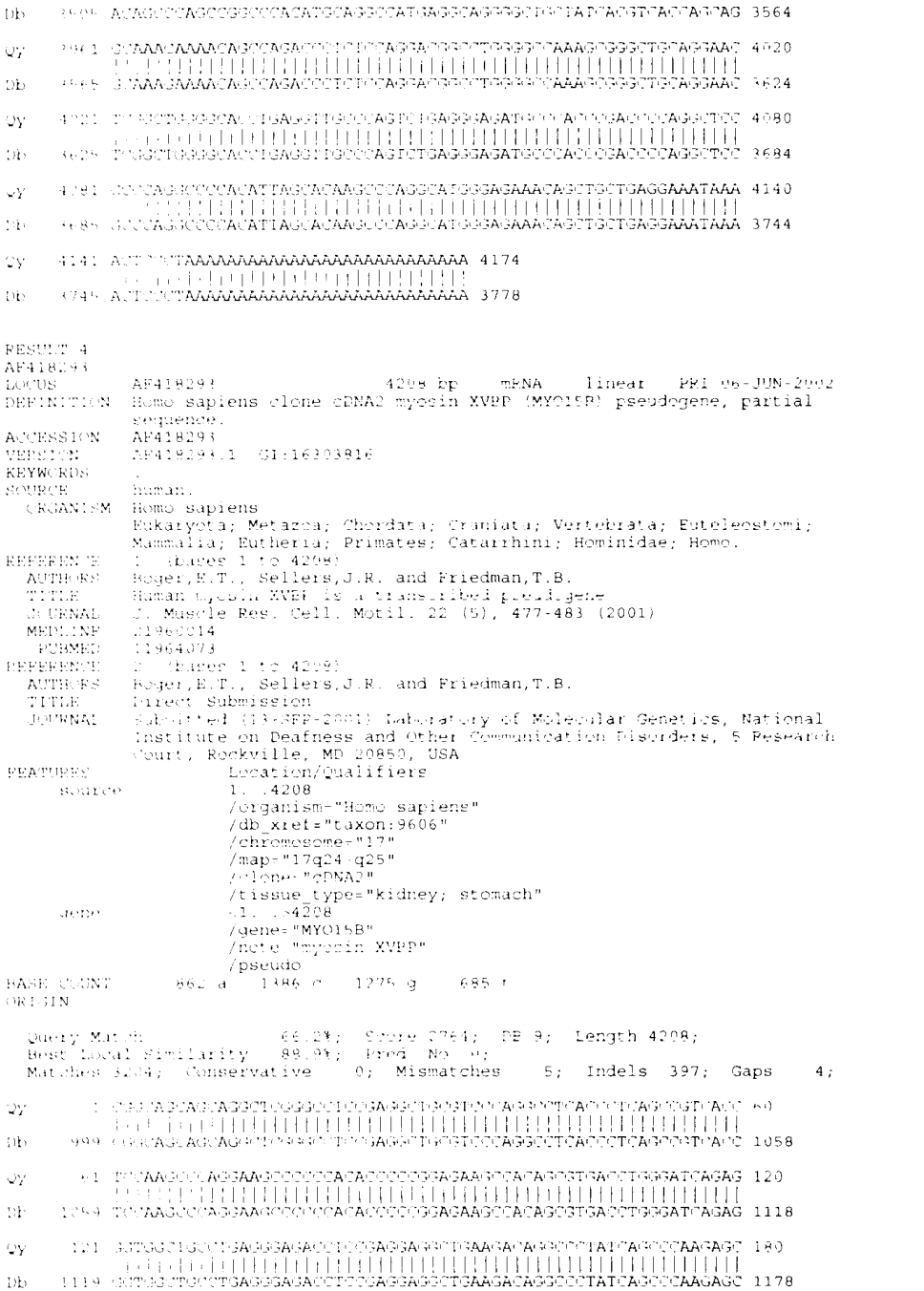
Qy       1501 GGGCCATCGAGGGCGCTGGTTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATG 1560
Db       1501 GGGCCATCGAGGGCGCTGGTTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATG 1560

Qy       1561 TCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCA[illegible]CACCGTGGGG 1620
Db       1561 TCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCA[illegible]CACCGTGGGG 1620

Qy       1621 ACCTCATCAAGCTGCTGCCGGTGGCCACCCTGGAGCCAGG[illegible] 1680
Db       1621 ACCTCATCAAGCTGCTGCCGGTGGCCACCCTGGAGCCAGG[illegible] 1680

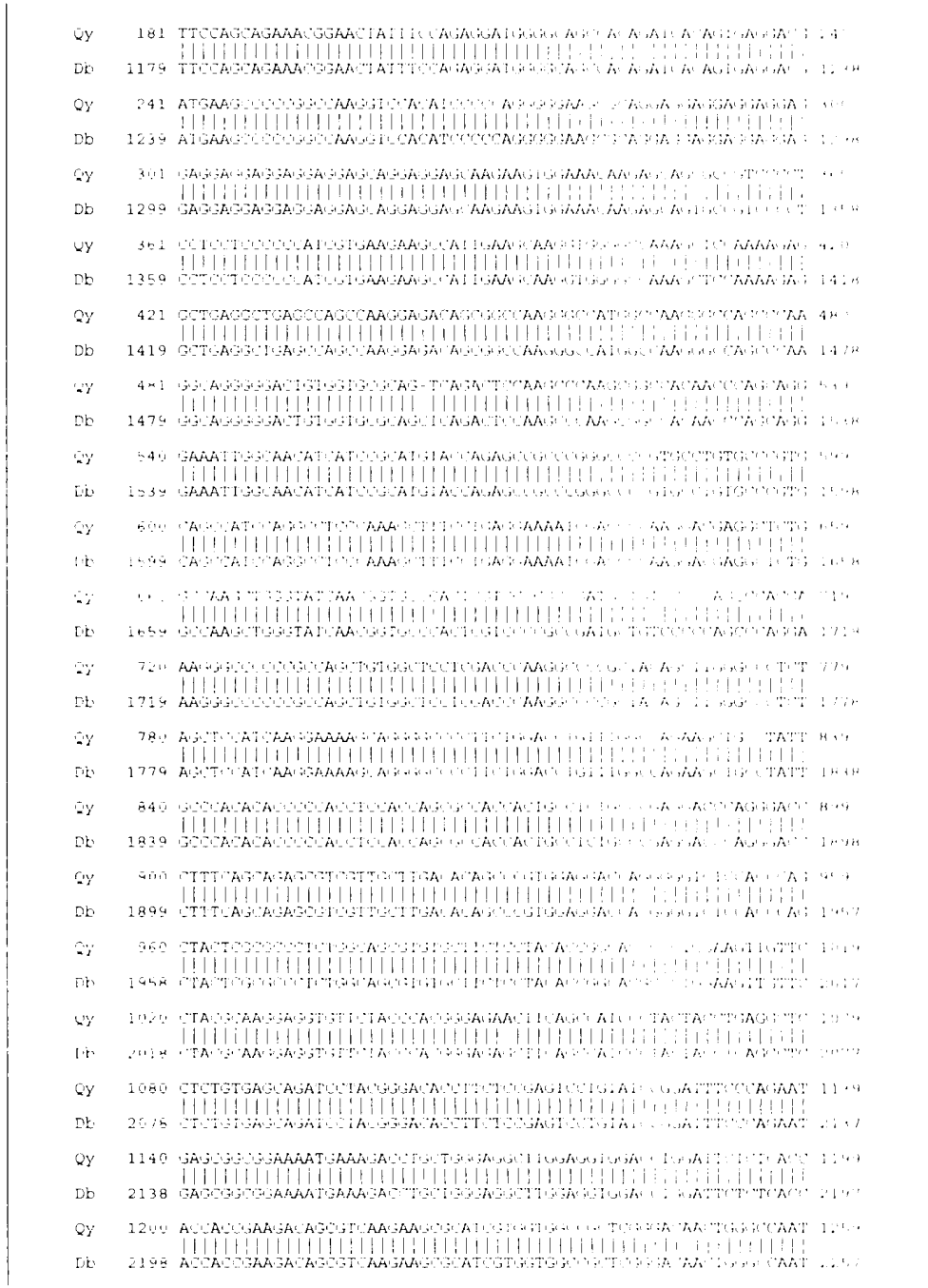
Qy       1681 GGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTGCAGCCGC[illegible]GACTTTT 1740
Db       1681 GGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTGCAGCCGC[illegible]GACTTTT 1740
```

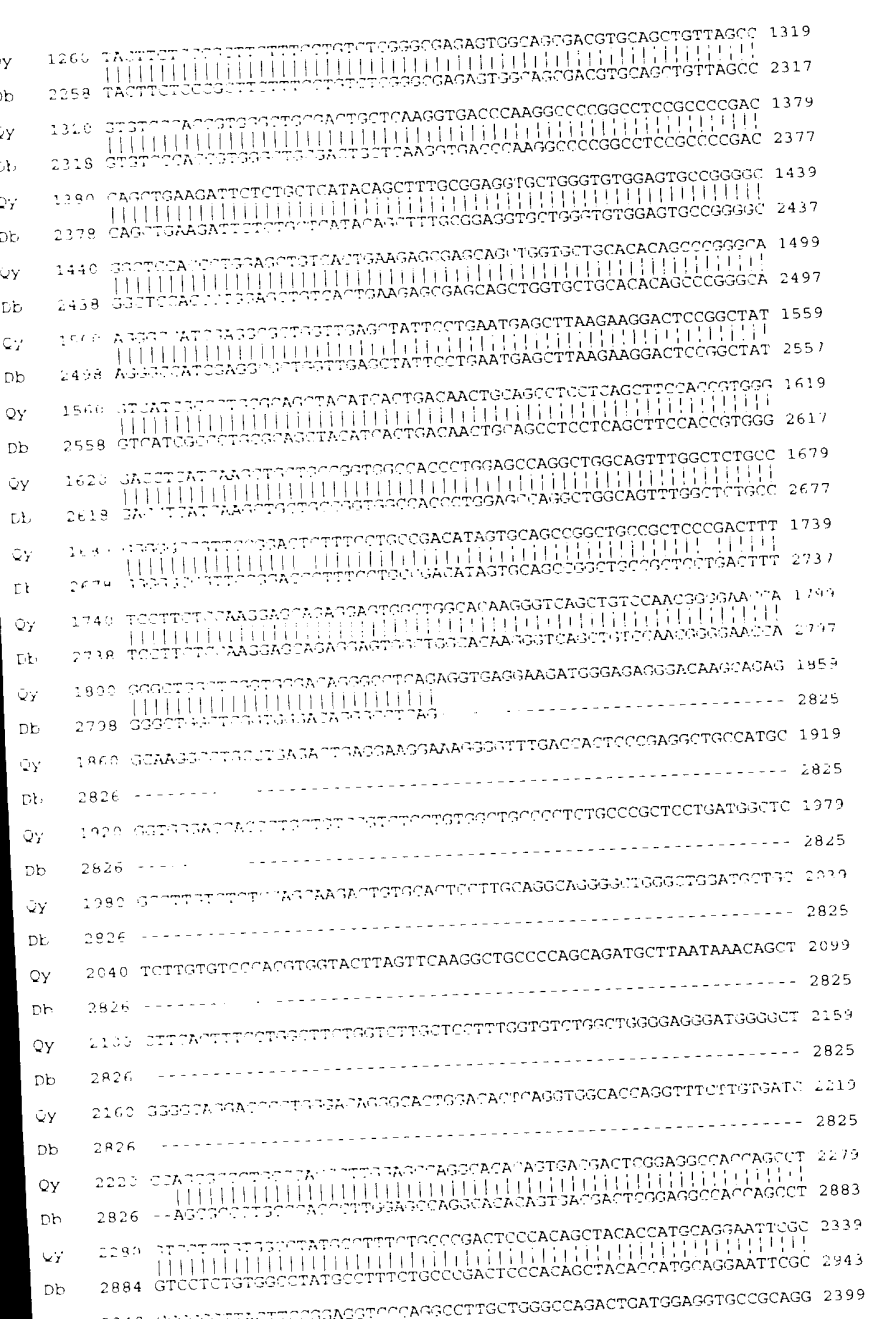


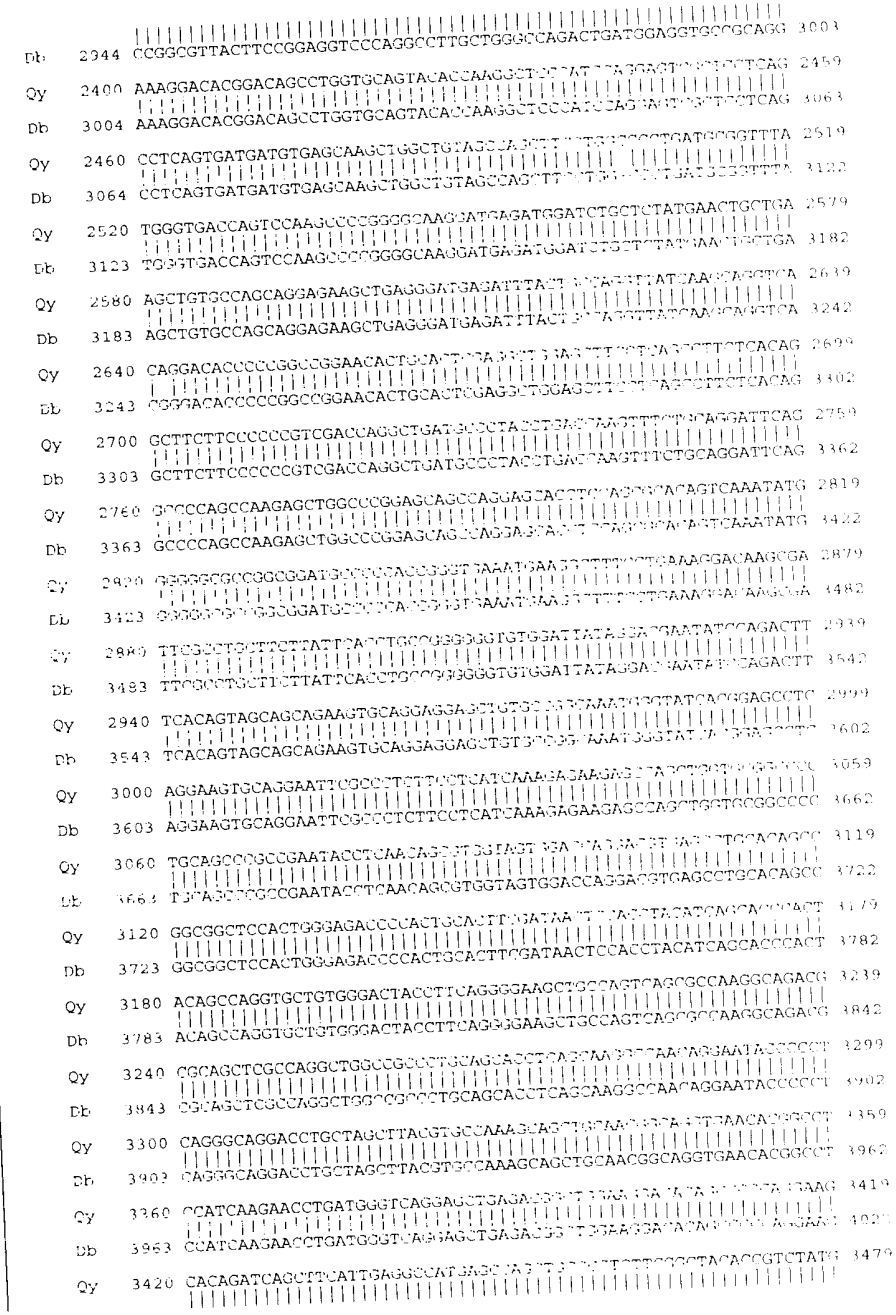


```
RESULT 4
AF418293
LOCUS       AF418293                4208 bp    mRNA    linear   PRI 06-JUN-2002
DEFINITION  Homo sapiens clone cDNA2 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418293
VERSION     AF418293.1  GI:16303816
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4208)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21960014
    PUBMED  11964073
REFERENCE   2  (bases 1 to 4208)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..4208
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="cDNA2"
                     /tissue_type="kidney; stomach"
     gene            <1..>4208
                     /gene="MYO15B"
                     /note "myosin XVBP"
                     /pseudo
BASE COUNT      862 a   1486 c   1275 g    685 t
ORIGIN

  Query Match             66.2%;  Score 2764;  DB 9;  Length 4208;
  Best Local Similarity   88.0%;  Pred. No. 0;
  Matches 3206;  Conservative    0;  Mismatches    5;  Indels  397;  Gaps    4;
```







```
Db    4024 CATAGATTAGCTTCATTGAGGCCATGAGCCAGCTGCCCTTCTTCTGGCTACACTCTATG 4083

Qy    3480 GGGTGCTGCGAGTGAGCATGCAAGCCCTCTTCGGACCCACTCTCCTCGGGCTCAAACCCC 3539

Db    4084 GGATGCTCCGAGTGAGCATGCAGGCCCTCTTCGGACCCACTCTCCTCGGGCTCAAACCCC 4143

Qy    3540 AGTATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGGCCCTGAAGA 3599

Db    4144 AGTATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGGCCCTGAAGA 4203

Qy    3600 GCTGC 3605

Db    4204 GCTGC 4208


RESULT 1
BC027875
LOCUS       BC027875                2347 bp    mRNA    linear   PRI 01 MAY 2002
DEFINITION  Homo sapiens, hypothetical protein FLJ22686, clone MGC:34148
            IMAGE:5205535, mRNA, complete cds.
ACCESSION   BC027875
VERSION     BC027875.1  GI:20379900
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2347)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08 APR 2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC Help Desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,F., Benkert,M. Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brooks,S., Coleman,H.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Laric,P., Legaspi,R., Maduro,Q.L.,
            Marfella,C., Masiello,P., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C.,
            Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A.,
            Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 49 Row: g Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 13376444.
FEATURES    Location/Qualifiers
     source 1..2347
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="MGC:34148 IMAGE:5205535"
            /tissue_type="Lung, Spleen, fetal, pooled"
            /clone_lib="NIH_MGC_122"
            /lab_host="DH10B"
            /note="Vector: pCMV-SPORT6"
     CDS    418..1881
            /codon_start=1
            /product="hypothetical protein FLJ22686"
            /protein_id="AAH27875.1"
            /db_xref="GI:20379901"
            /translation="MQEPARRYFPPSQALLSQTSPGAAGKITDSLVQYTPAPCQPSLL
            ...SSTVCFLAVAGFLALMPFMGQ...SRFSSCLYTLLPFL...FFLPFFSPCVTP
            QVTGHPRPEHCTRGWSFLSLLTGFFPPSTPLMPYLTKPTQLDSGSGQELARGSLRRL...P
            TVPYGTSPPMFTPGMPAPCPPQAGSLSTPGT...TPTSAAPC...PPL...PG
            MGITEPQPEVQEFALFLIKEPSSQLVRPLQPAEYL...VVVLQLVSLDSRRVHWPTDLRPL
            NGTYTSTHTSQVLWDYLQGKLLVGAKADAQLAKLAALQHLGKANKNTPGG...LLAYYL
            KQLQRQVNTAGIKNLMGQELRRLELGHSPQEA...IGPTIAMSQLLDLIYTVYVVLNVSM...
            AIGGPTL...GLNPQHGTLMTP...PQPLTPPIAL...PLPLPPGSTLTPPPPSSA
            DNPQTIWPELPQAQELLYTTVPLIDSSASGTEWPSIN"
BASE COUNT      574 a    719 c    654 g    400 t
ORIGIN

  Query Match           53.8%;  Score 2246;  DB 9;  Length 2347;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 2260;  Conservative    0,  Mismatches    5;  Indels    1;  Gaps    1;

Qy    1909 GGCTGCCATGCGGTGGGACCACCTGCTGTCCCTTCTCCTCTGCCCCCT 1958
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGCTGCCATGCGGTGGGACCACCTGCTGTCCCTTCTCCTCTGCCCCCT 50

Qy    1959 CCTGATGGGCTGGCCTTGTCTCTCCAGCAAGACTGTGCACTCTCTGCAGCAAGCAGTGGT 2018
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 CCTGATGGGCTGGCCTTGTCTCTCCAGCAAGACTGTGCACTCTCTGCAGCAAGCAGTGGT 110

Qy    2019 CTGGATGCTGCTCTTGTGTCCCACGTGGTACTTAGTTCAAGCAGAGATCTTT 2068
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 CTGGATGCTGCTCTTGTGTCCCACGTGGTACTTAGTTCAAGCAGAGATCTTT 160

Qy    2069 AATAAACAGCTCTTCACTTTGCTGGCTTCTGGTCTTGCTCTTTGGTGTCTGCTTGGGA 2128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 AATAAACAGCTCTTCACTTTGCTGGCTTCTGGTCTTGCTCTTTGGTGTCTGCTTGGGA 220

Qy    2129 GGGATGGGGCTGGGGCAGGACCCCTGGGACAGGGCACTGGACACTCAGGTGCACTCAGGT 2188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GGGATGGGGCTGGGGCAGGACCCCTGGGACAGGGCACTGGACACTCAGGTGCACTCAGGT 280

Qy    2189 TTCTTGTGACTCCAGGCCCTGCCACCCTGGAGCCAGGACACAGTCACAGTGACTCTGGAG 2248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 TTCTTGTGACTCCAGGCCCTGCCACCCTGGAGCCAGGACACAGTCACAGTGACTCTGGAG 340

Qy    2249 GCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCACACGTACAGCTATC 2308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 GCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCACACGTACAGCTATC 400

Qy    2309 CAGGAATTCCCCCCGGGTTACTTCCTGGAAGTCCCAGGCCTTCCTGGGCCAGACTGACTGA 2368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 CAGGAATTCCCCCCGGGTTACTTCCTGGAAGTCCCAGGCCTTCCTGGGCCAGACTGACTGA 460

Qy    2369 GGTGCCGCAGGAAAGGACAGGACAGCTGAGTGCAGTACACCAAGGTTCCTATCCAGAG 2428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 GGTGCCGCAGGAAAGGACAGGACAGCTGAGTGCAGTACACCAAGGTTCCTATCCAGAG 520

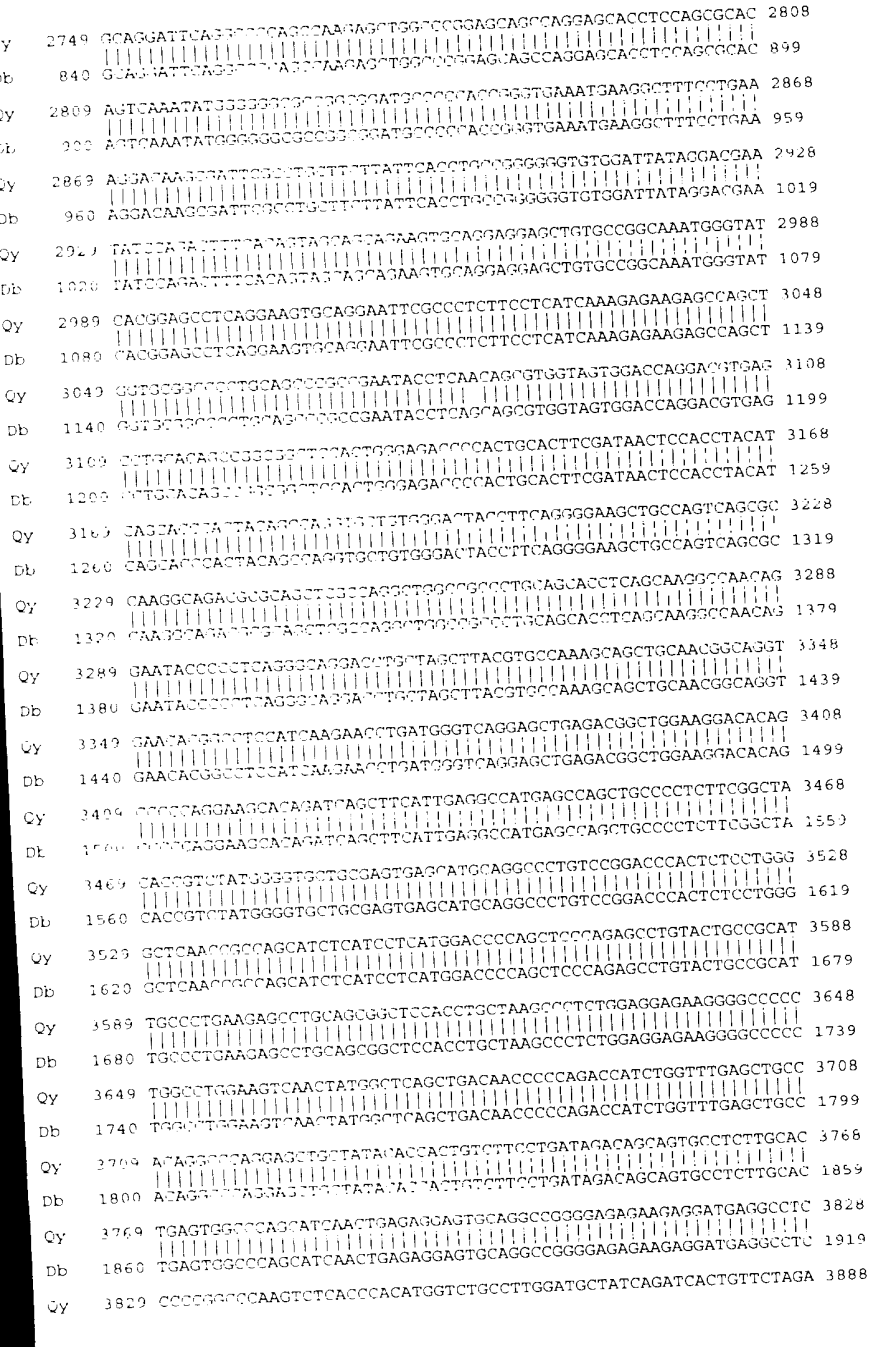
Qy    2429 TGGCTCCTCAGGCTCAGTGATGATGTGAGCAAGTGGGTCATCCTGTTTTAGGCCCCT 2488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 TGGCTCCTCAGGCTCAGTGATGATGTGAGCAAGTGGGTCATCCTGTTTTAGGCCCCT 580

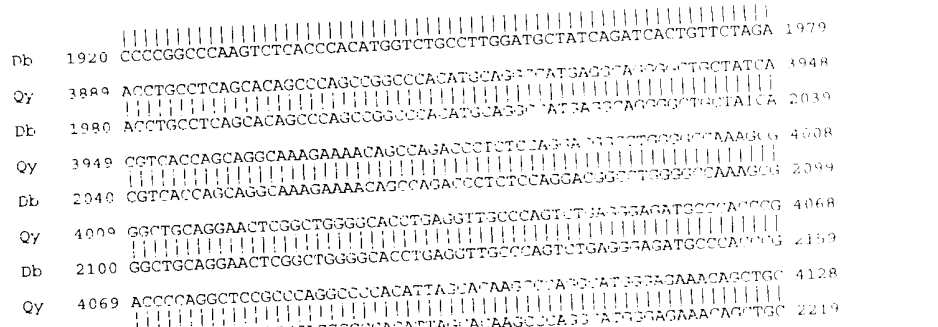
Qy    2489 GATGGGGTTTATGGGTGACTACTCTCAAGGCCAAGAGCAGAGAGAGATGTTA 2548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 GATGGGGTTTATGGGTGACTACTCTCAAGGCCAAGAGCAGAGAGAGATGTTA 640

Qy    2549 TGAACTGCTGAAGCTGTGCCAGCAGGACAAGCTGAGGGATGACATTTACTTGCAGCTTAT 2608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 TGAACTGCTGAAGCTGTGCCAGCAGGACAAGCTGAGGGATGACATTTACTTGCAGCTTAT 700

Qy    2609 CAAGCAGTCTCAGAGAGAGAGGAAGAGAACAGAGTGAGGAAGCAGCTTCAGAAGTTAG 2668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     701 CAAGCAGGTCACGGAGAGAGAGGAAGAACAGTGCACTCGAGGCCGGAGCTCCTAG 760

Qy    2669 CCTTCTCACAGGCTTCTTCCCCCGTCGACCAGCTGATGCCTACTGACAAGTTTT.T 2728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     761 CCTTCTCACAGGCTTCTTCCCCCGTCGACCAGCTGATGCCTACTGACAAGTTTTCT 820
```





```
RESULT 6
AB058686
LOCUS       AB058686                6004 bp    mRNA    linear   PRI 05-JUN-2001
DEFINITION  Homo sapiens mRNA for KIAA1783 protein, partial cds.
ACCESSION   AB058686
VERSION     AB058686.1  GI:14017782
KEYWORDS
SOURCE      Homo sapiens brain cDNA to mRNA, clone: fg01285.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. XX.
            The complete sequences of 100 new cDNA clones from brain which code
            for large proteins in vitro
  JOURNAL   DNA Res. 8 (2), 85-95 (2001)
  MEDLINE   21245130
REFERENCE   2  (bases 1 to 6004)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:dnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..6004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="fg01285"
                     /tissue_type="brain"
                     /note="vector:pBluescriptII SK plus"
     gene            1..6004
                     /gene="KIAA1783"
     CDS             <1081..5763
                     /gene="KIAA1783"
                     /note="Start codon is not identified."
                     /codon_start=1
                     /product="KIAA1783 protein"
                     /protein_id="BAB47412.1"
                     /db_xref="GI:14017783"
                     /translation="PVCCAGSFRLDPVGYTASALGITHPCALARMKEAGVVRRGHSPM
                     CPFHCHMPPGPAGFWPGATVQQPRQRAPTALLQPLAAGPGGSVCAGVAVVGACPSAS
                     EGVLPFFPGPSAPQPPEYPGRPDMAVPGWSAMVRSRLTATSTCQATDPPPLQPSHTHH
                     GDHPSYAKPRLFLPVFTVPHYAGTVTYQVHFFLMFNPDQLPPAVVFMLGQSQLQLVGS
                     LFQEAEPQSRGGRGPRTLASPFQQALEDLIARLGRSHVYFIQCLTFMLQELIQEVG
                     HVTEQLHQAAILEAVGTPSANFFVRVPFEAFLASPQALGSEGQELLGPPEPCGAVLSQ
                     VLGAESPLYHLGATKVLLQEQGWQKLEELKDQQRSQALVDLHRESHTTQRQRVLERM
```

QARMRGPQARPPYLPPPAALGQINTILIVAQPLLQPRQPLQLGPWQGWHSSPRALERV
PLMLLGRLEILALLAVMLRTAESHRIALAGSITELLPEVPAAPSLTLPADILLFPP
SPVAIGPQEPSLPRPGQPLAAPLPLQLDGDNPQRALDINKVMLRLLGDGGLESWQRQIM
GAYLVRQGQCWPGLRNELFSQLVAQLWQNIDEQQSQKGWALMAVLLSAPEPLPVLQKP
LLRFVSDQAPRGMAALQQHKLLGALEQSQLASGATPAHPPTQLEWLAGWPRGPMALDV
FTFSEECYSAEVESWITGELAGWILQSPGLEAPPPGWSVSLHSRDAWLQLAPCDFVI
DLISQTEDLGDPARPPSYPITPIGSAFAIPLAPGIQAPSLPPGPPPGPAPTLPSPDHT
GEVQRSGSLIQSPLDQIFLPVISSGLSPLRQSWALSSRMPGGGATGPTQQSYPMVYPGM
IQMFAYQPGMVFAFMPMMPAMGTVFAMPAMVVPFQPPLPSLDAPQLAVSQQNFIQQQA
LILAQQMTAQAMSISIPQQMQLPQQLAPASEAASQLASPSAVTSKPRKPPTPPEPRQD
LGSEGGCLRETSEEAEDRPPYQPPSPQQPRPNYPQPMGQPQITVPTMKPPAKVHIPQGEA
QEEEEEEEEEEEQEEQEVETRAVPSPPPPPTVKKPLPQQGAKAPPPAPAEPAPETAAP
GHGQGPAQGPGTVVPSSTSKPKKPQPSPEIGNIIPMYQSPPGPVPVPVQPSPPPKAPL
RKIEPPLEALAPYDNQAHSSFMLSPPPGPGPFPTAVAIPPPAPLQIGPPSSIPEPQS
PLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERPCLTQPVEDQGVSTQLLAPSG
SVCFSYTGIPWPPLPPEVFYPFNPPHPTTISLLPQLLKQTFSEFITLQQHEPF
MKDLLGGLEVDLDSLITTEDSVKKPIVVAARDNWANYFSRFFPVSGESGSDVQLLAVS
HPGIPLLFVTQTPGLRFDQLPIICSYSFAEVLGVEQPGDSTLELNLFSELVLHTARA
RAIEALVELFDNELPKDSGYVIALPSYITDNCSLLSFHPGDLIKILPVATLEPGWQFG
SAGGPSGLFPADIVQPAAAKPSPSFEQPSGWHPGQLSNTEPSLAPWDPASPVPPMGE
GQAEARPA"

BASE COUNT 1134 a 1942 c 1870 g 1058 t
ORIGIN

Query Match 50.4%; Score 2102.4; DB 9; Length 6004;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2114; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy       1 CCCCAGCAGCAGGCTCCGCCCCTCGGAGGCTGCGTCCCAGGCCTCAGCCTCAGTCCTCAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3949 CCCCAGCAGCAGGCTCCGCCCCTCGGAGGCTGCGTCCCAGGCCTCAGCCTCAGTCCTCAGC 3948

Qy      61 TCCAAGCCCAGGAAGCCCCCGCAGACTCCCCGGAGAAGCCTACAGCGTGACCTGGGATCAGAG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3948 TCCAAGCCCAGGAAGCCCCCGCAGACTCCCCGGAGAAGCCTACAGCGTGACCTGGGATCAGAG 4008

Qy     121 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAAGACAGGCCCTATCAGCCCAAGAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4009 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAAGACAGGCCCTATCAGCCCAAGAGC 4068

Qy     181 TTTCAGCAGAAACGCAACTATTTCCAGAGGATGGGGCTAGCCACAGATCACAGTGAGGACG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4069 TTTCAGCAGAAACGCAACTATTTCCAGAGGATGGGGCTAGCCACAGATCACAGTGAGGACG 4128

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4129 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAGAG 4188

Qy     301 GATCAGGAGGAGGAGGAGGACGAGGAGGAGCAAGAAGTGGAAACAAGAGCAGTGGCGTCCCCT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4189 GATCAGGAGGAGGAGGAGGAGGAGGAGGAGCAAGAAGTGGAAACAAGAGCAGTGGCGTCCCCT 4248

Qy     361 CTTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAAGCTCCAAAAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4249 CTTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAAGCTCCAAAAGAG 4308

Qy     421 GCCAGGAAGGAGCCACCAAGGAGACAGTGGGCTAAGGGCATGGCCAAGGGCCAGCCCAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4309 GCCAGGAAGGAGCCACCAAGGAGACAGTGGGCTAAGGGCATGGCCAAGGGCCAGCCCAA 4368

Qy     481 GCCAGGGGGACTGTGGTGCGCAGCTCACACTCCAAGCCCAAGTGGCCACAACCCAGTAGG 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4369 GCCAGGGGGACTGTGGTGCGCAGCTCACACTCCAAGCCCAAGTGGCCACAACCCAGTAGG 4428

Qy     540 GAAATTGCCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCTG 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4429 GAAATTGCCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCTG 4488

Qy     600 AAGCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTG 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4489 CAGCTATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTG 4548

Qy     660 GCCAAGCTGGGCATCAAGGTGGTGCCACTCGTGCCCGGCGATGCTGTCCCCCAGCCCAGGA 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4549 GCCAAGCTGGGCATCAAGGTGGTGCCACTCGTGCCCGGCGATGCTGTCCCCCAGCCCAGGA 4608

Qy     720 AAGGGGCCCCCGGCCAGCTCTGTGGCTCCCGGACCCAAGCCCCCCCCCCTACCTTCT 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4609 AAGGGGCCCCCGGCCAGCTCTGTGGCTCCCGGACCCAAGCCCCCCCCCTACCTTCT 4668

Qy     780 ACTCTCACTCAAAGAAAAAGAAGGAAACCCTTCCGGACCCGTTTTGCCCAGAAGCTCTCTATT 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4669 ACTCTCACTCAAAGAAAAAGAAGGAAACCCTTCCGGACCCGTTTTGCCCAGAAGCTCTCTATT 4728

Qy     840 CCCCACACACCCCCACCCCCACCCCACCCACCCACCCCCCCCCCCCCCCCCCCCCCCCCCC 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4729 CCCCACACACCCCCACCCCCACCCCACCCACCCACCCCCCCCCCCCCCCCCCCCCCCCCCC 4788

Qy     900 CTTTCACCAGACAGCCTCCTTCTTCAACACCCCTCCACACCACCCCCCCCCCCCCCCCCCC 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4789 CTTTCACCAGACAGCCTCCTTCTTCAACACCCCTCCACACCACCCCCCCCCCCCCCCCCCC 4848

Qy     960 CTACTCCGCCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACGCCCTCTGGAAGTTCTTC 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4849 CTACTCCGCCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACGCCCTCTGGAAGTTCTTC 4908

Qy    1020 CTACGCAAGGAGGGGTCTCTACCCACCCGGAGAACTTCAGCCATCCCTACTACCTGAGGCTC 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4909 CTACGCAAGGAGGGGTCTCTACCCACCCGGAGAACTTCAGCCATCCCTACTACCTGAGGCTC 4968

Qy    1080 CTCTGTGAGGCAGATCCTACGGGACACCTTCTCCGAGTCCCTGATCCAGAATTTCCTAGAAT 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4969 CTCTGTGAGGCAGATCCTACGGGACACCTTCTCCGAGTCCCTGATCCAGAATTTCCTAGAAT 5028

Qy    1140 GAGCGGCGGAAAATGAAAGACCTGCTGGGAGGGCTTTGAGGTCTATGTGCTGACCTCAGCTT 1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5029 GAGCGGCGGAAAATGAAAGACCTGCTGGGAGGGCTTTGAGGTCTATGTGCTGACCTCAGCTT 5088

Qy    1200 ACCACCGAAAGACAGGTCCAAGAAGCGGCATCGTGGTGGTGGCCGTGGAGACAATTCCCAAT 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5089 ACCACCGAAAGACAGGTCCAAGAAGCGGCATCGTGGTGGTGGCCGTGGAGACAATTCCCAAT 5148

Qy    1260 TACTTCTCCCGCTTCTTTCCTGTCTCTGGGCGAGAGTGGCAGCTACTCTCAGTTGTTAGAT 1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5149 TACTTCTCCCGCTTCTTTCCTGTCTCTGGGCGAGAGTGGCAGCTACTCTCAGTTGTTAGAT 5208

Qy    1320 GTGTCCCACCGTGGGCTGGCACTGCTCAAGGTGACCCAAGGCACCCCCCGGACTCTGCCCGAC 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5209 GTGTCCCACCGTGGGCTGGCACTGCTCAAGGTGACCCAAGGCACCCCCCGGACTCTGCCCGAC 5268

Qy    1380 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGAGTGCTGGCTGCGACGGCAAGGAC 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5269 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGAGTGCTGGCTGCGACGGCAAGGAC 5328

Qy    1440 GGCTCCACCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGGTGACACAGCCCGGCTA 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5329 GGCTCCACCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGGTGACACAGCCCGGCTA 5388

Qy    1500 AGGGCCATCGAGGGCGTCTTCGTGAGCTACTTCCTGAATGAGCTTAAGAAGCTACTCGGCTAT 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5389 AGGGCCATCGAGGGCGTCTTCGTGAGCTACTTCCTGAATGAGCTTAAGAAGCTACTCGGCTAT 5448

Qy    1560 GTCATCGCCCTGGCCAGCTACATCACTGACAACTTCGAGCTCCTCAGCATCCCCACCGTGGG 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5449 GTCATCGCCCTGGCCAGCTACATCACTGACAACTTCGAGCCTCCTCAGTTTCCACCGTGGG 5508

Qy    1620 GACCTCATCAAGCTGCTGCCGGTGGCCAGCCTGGAGCCAGGCTGCACGTTTGCCTCTGGC 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5509 GACCTCATCAAGCTGCTGCCGGTGGCCAGCCTGGAGCCAGGCTGCACGTTTGCCTCTGGC 5568

Qy    1680 GGGGGCCGTTCCGGACTCTTTCCTGGCGACATACTGCAGCCGGCTGCCGACTTT 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5569 GGGGGCCGTTCCGGACTCTTTCCTGGCGACATACTGCAGCCGGCTGCCGACTTT 5628

Qy    1740 TCCTTCTCCAAGAAGACACAGGAGTGGTTGCAACAAGTTACATCTGTCTCAACTCCGAACCA 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    5629 TCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTCAGCTGTCCAACGGGGAACCA 5688

Qy    1800 GGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGGAAGATGGGAGAGGGACAAGCAGAG 1859

Db    5689 GGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGGAAGATGGGAGAGGGACAAGCAGAG 5748

Qy    1860 GCAAGGGCTGCCTGACACTGAGGAAGCAAAGGGTTTGACCACTCCCGAGGCTGCCATGC 1919

Db    5749 GCAAGGGCTGCCTGACACTGAGGAAGCAAAGGGTTTGACCACTCCCGAGGCTGCCATGC 5808

Qy    1920 GGTGGGACCACCCTGCCCTCCCTGTGGCTGCCCCTCTGCCCGCTCCTGATGGCTC 1979

Db    5809 GGTGGGACCACCCTGCCCTCCCTGTGGCTGCCCCTCTGCCCGCTCCTGATGGCTC 5868

Qy    1980 GCCTTGTCTCTCCAGCAAGACTGTGCACTCCTTGCAGGCAGGGGCTGGGCTGGATGCTGC 2039

Db    5869 GCCTTGTCTCTCCAGCAAGACTGTGCACTCCTTGCAGGCAGGGGCTGGGCTGGATGCTGC 5928

Qy    2040 TCTTGTGTCCCACGTGGTACTTAGTTCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCT 2099

Db    5929 TCTTGTGTCCCACGTGGTACTTAGTTCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCT 5988

Qy    2100 CTTCACTTTCCTGGCT 2115

Db    5989 CTTCACTTTCCTGGCT 6004
```

RESULT 7
AK097896
LOCUS AK097896 2372 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ40577 fis, clone THYMU2007467, weakly similar to Mus musculus unconventional myosin-15 mRNA.

ACCESSION AK097896
VERSION AK097896.1 GI:21757796
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens shymus cDNA to mRNA, clone_lib:THYMU2 clone:THYMU2007467.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,F., Arita,M., Musashino,E., Yuuki,H., Hara,H., Sugiyama,T., Irie,R., Suzuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2372)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438 52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan, cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.), 5' & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES Location/Qualifiers
source 1..2372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYMU2007467"
/tissue_type="shymus"
/clone_lib="THYMU2"
/note="cloning vector: pME18SFL3"

BASE COUNT 486 a 780 c 715 g 391 t
ORIGIN

Query Match 50.0%; Score 2098.6; DB 9; Length 2372;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2105; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```
Qy       1 CGGCAGCAGCAGGCTCGGGGCTCGGAGGCTGGGTGCCAAGGCCTCAGCCGTCACC 56

Db     258 CGGCAGCAGCAGGCTCGGGGCTCGGAGGCTGGGTGCCAAGGCCTCAGCCGTCACC 313

Qy      57 TCCAAGCCCAGGAAGGCCCCACACCCCGGAGAAGCCACAGGCTGACCTGGGATCAGAG 116

Db     314 TCCAAGCCCAGGAAGGCCCCACACCCCGGAGAAGCCACAGGCTGACCTGGGATCAGAG 373

Qy     117 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCTATCAGCCCAAGAGC 176

Db     374 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCTATCAGCCCAAGAGC 433

Qy     177 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 236

Db     434 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 493

Qy     237 ATGAAGCCCCCGGCCAAGGTCCACATCCCCAGGGGGAAGCCGCAGGAGGAGGAGGAGGAG 296

Db     494 ATGAAGCCCCCGGCCAAGGTCCACATCCCCAGGGGGAAGCCGCAGGAGGAGGAGGAGGAG 553

Qy     297 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCCCCGTCCCCT 356

Db     554 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCCCCGTCCCCT 613

Qy     357 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 416

Db     614 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 673

Qy     417 GCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 476

Db     674 GCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 733

Qy     477 GGCAGGGGGACTGTGGTGCGCAG-TCAGACTCCAAGCCCAAGCCGCCACAACCCAGCAGG 535

Db     734 GGCAGGGGGACTGTGGTGCGCAGCTCAGACTCCAAGCCCAAGCCGCCACAACCCAGCAGG 793

Qy     536 GAAATTGGCAACATCATCCGCATGTACCAGAGCCGCCGGGGCCCCGTGCCTGTGCCCGTG 595

Db     794 GAAATTGGCAACATCATCCGCATGTACCAGAGCCGCCGGGGCCCCGTGCCTGTGCCCGTG 853

Qy     596 CAGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTG 655

Db     854 CAGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTG 913

Qy     656 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGA 715

Db     914 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGA 973

Qy     716 AAGGGCCCCCCGCCAGCTGTGGCTCCTGGACCCAAGGCCGCCCTACAGCTTGGGCCCTCT 775

Db     974 AAGGGCCCCCCGCCAGCTGTGGCTCCTGGACCCAAGGCCGCCCTACAGCTTGGGCCCTCT 1033

Qy     776 AGCTCCATCAAGGAAAAGCAGGGGCCCCTTCTGGACCTGTTTGGCCAGAAGCCTGCCTATT 835

Db    1034 AGCTCCATCAAGGAAAAGCAGGGGCCCCTTCTGGACCTGTTTGGCCAGAAGCCTGCCTATT 1093

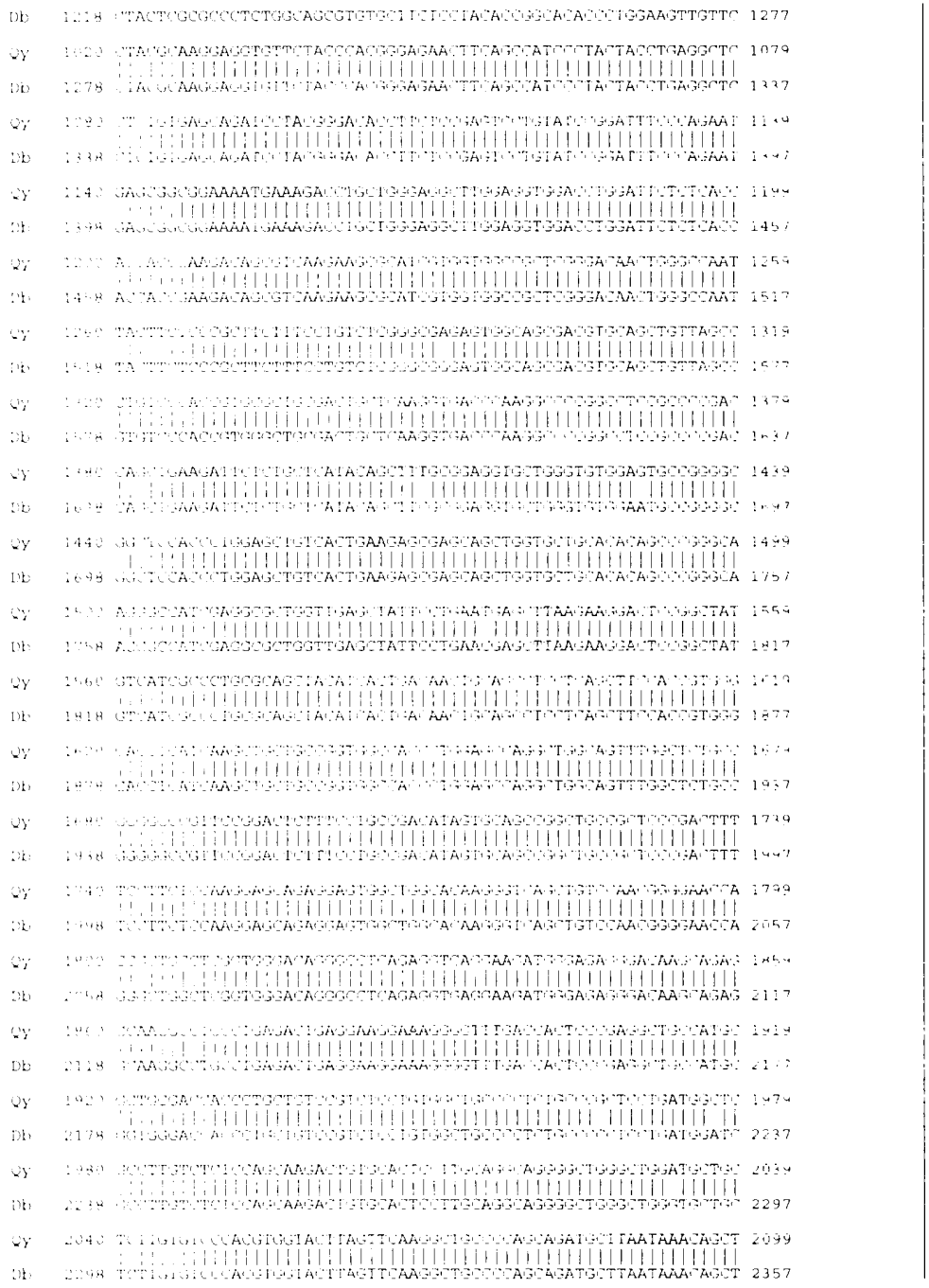
Qy     836 GCCCACACACCCCCACCTCCACCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACC 895

Db    1094 GCCCACACACCCCCACCTCCACCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACC 1153

Qy     896 CTTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAG 955

Db    1154 CTTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAG 1213

Qy     956 CTACTCGCGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACGCCCTGGAAGTTGTTC 1015
```

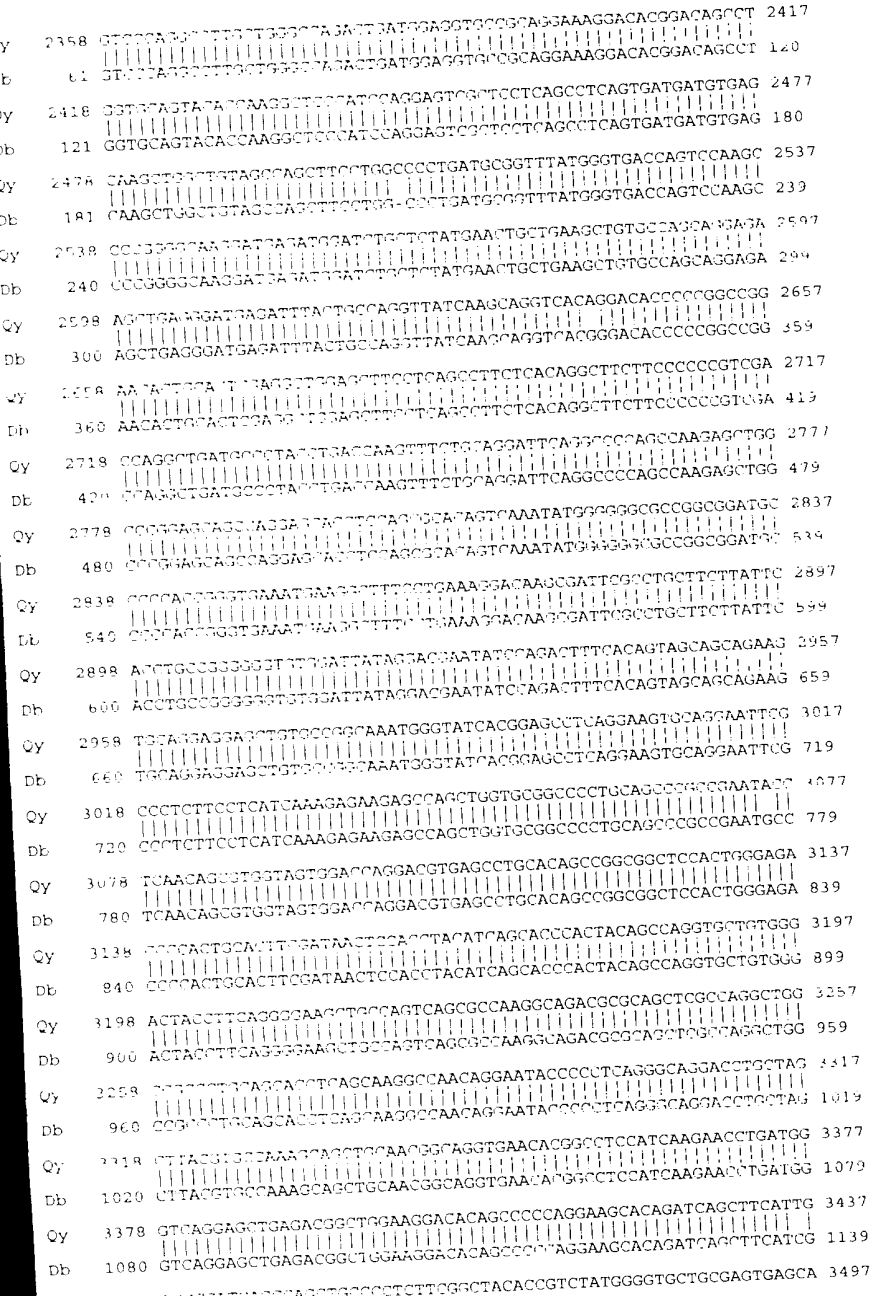


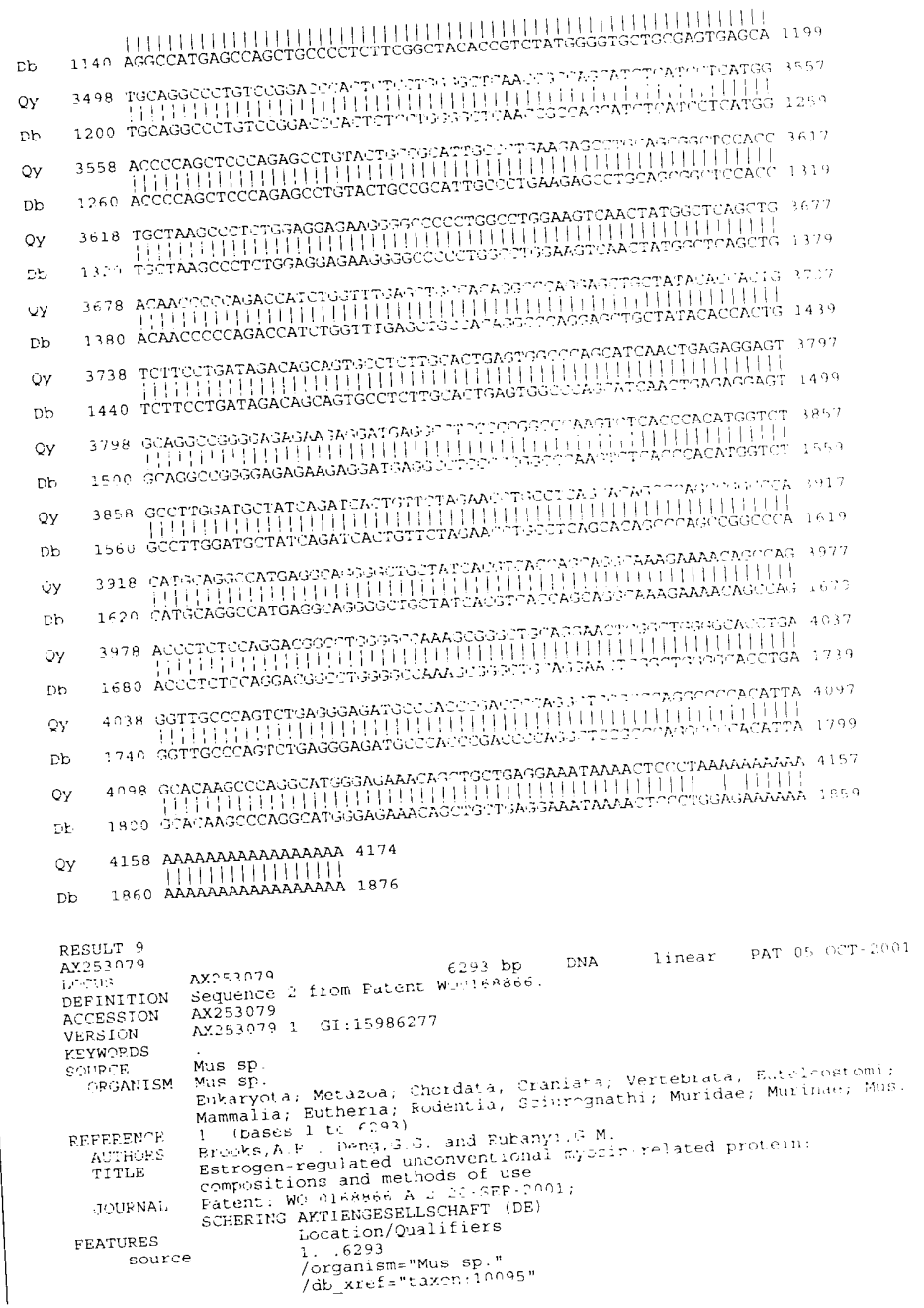
```
Qy   2100 CTTCACTTTCCTGGC 2114
          ||||||||||||||
Db   2358 CTTCACTTTCCTGGC 2372


RESULT 8
AK026339
LOCUS       AK026339               1891 bp    mRNA    linear   PRI 29-SEP-2002
DEFINITION  Homo sapiens cDNA: FLJ22686 fis, clone HSI10987.
ACCESSION   AK026339
VERSION     AK026339.1  GI:10439173
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens human small intestine cDNA to mRNA, clone lib:HSI
            clone:HSI10987.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Toshiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1891)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdnaadm@ims.u-tokyo.ac.jp, Tel:81-3-5449-5284,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Japanese and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..1891
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HSI10987"
                     /tissue_type="human small intestine"
                     /clone_lib="HSI"
                     /note="cloning vector pMF18SFL4"
     CDS             29..1492
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15453.1"
                     /db_xref="GI:10439174"
                     /translation="MQFFARRYFRRSQALLGQTDSGAAEKVTNSLVQYTKAPTQGGLL
                     SSSDVSFLAVAFLAMRPPMGPGSPLPGPPLPGPLPGVIP
                     QVTGHPRPPEHCTRGWSFLSLLTGFFPPSTRLMPYLTKFLQDSSPSQELARSSLPHLQP
                     TVEYGGRPPMPPPPSPMFAFLPGQAIRLLIHLIPSVLYWTNLQTFTVAAFVQPRLCPQ
                     MGITEPQFVQEFALPLIKEESQLVPPLQPAFCLNCYVVDQSVSLHSPRLHWFTPLHPL
                     NSSPISPTSQVLWSLPPVSAAADWLARLAACQHLSFANPNTPSQQPLAVVP
                     KQLQPQVNTASIPNLMSQELPELEGHSTQALLSPIEAMSQGQFLPGYTVYGVLRVSMQ
                     ALSGFTLGSLNNQMLIMGPSQGSLYCPLACPSQGRLLLTFFPGSTLPYVYSNA
                     DNPQTIWFELPQAQELLYTTVFLIDSSASCTEWPSIN"
BASE COUNT      464 a    536 c    536 g    306 t
ORIGIN

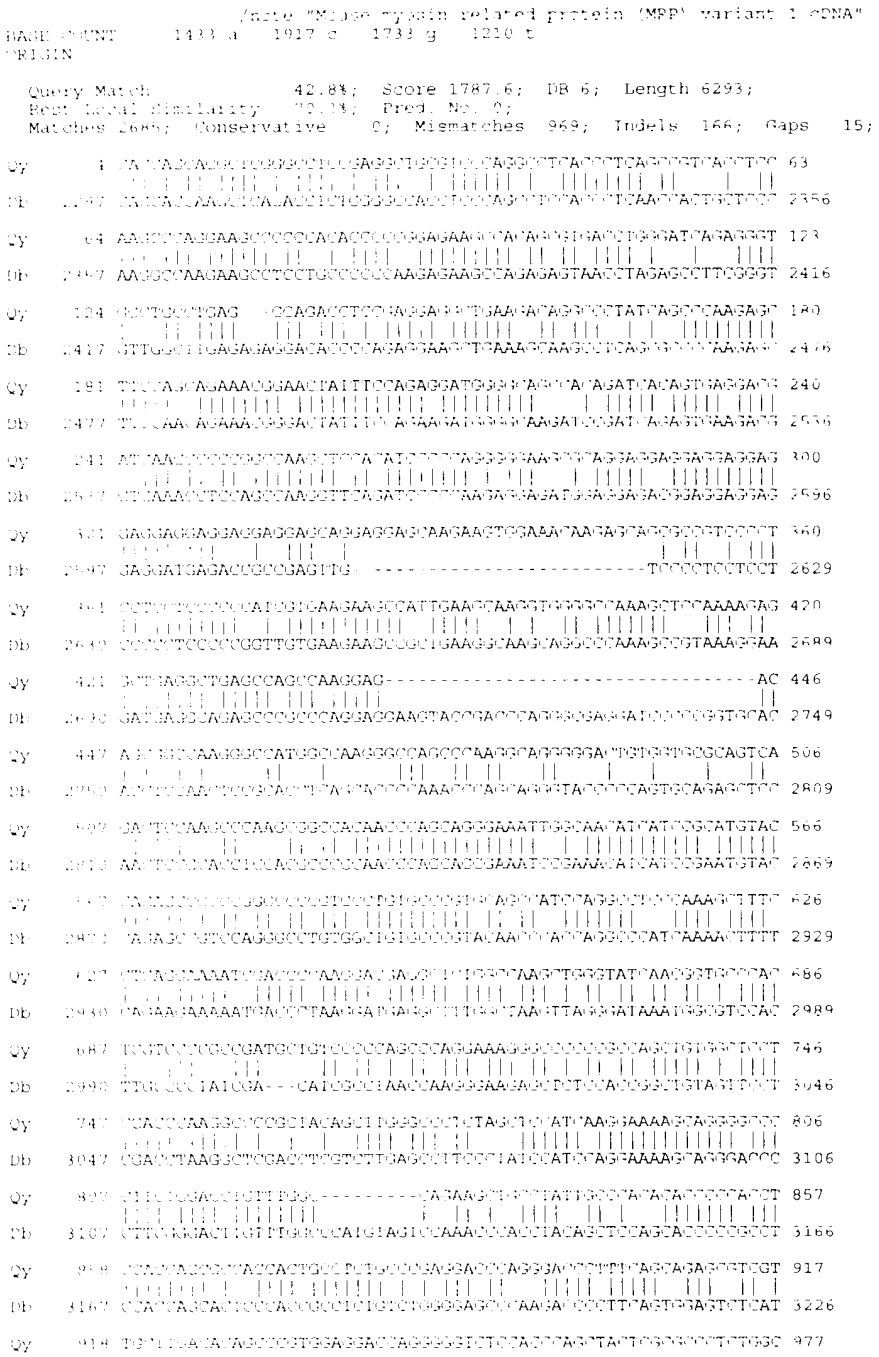
  Query Match             44.5%;  Score 1855.4;  DB 9;  Length 1891;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1870;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

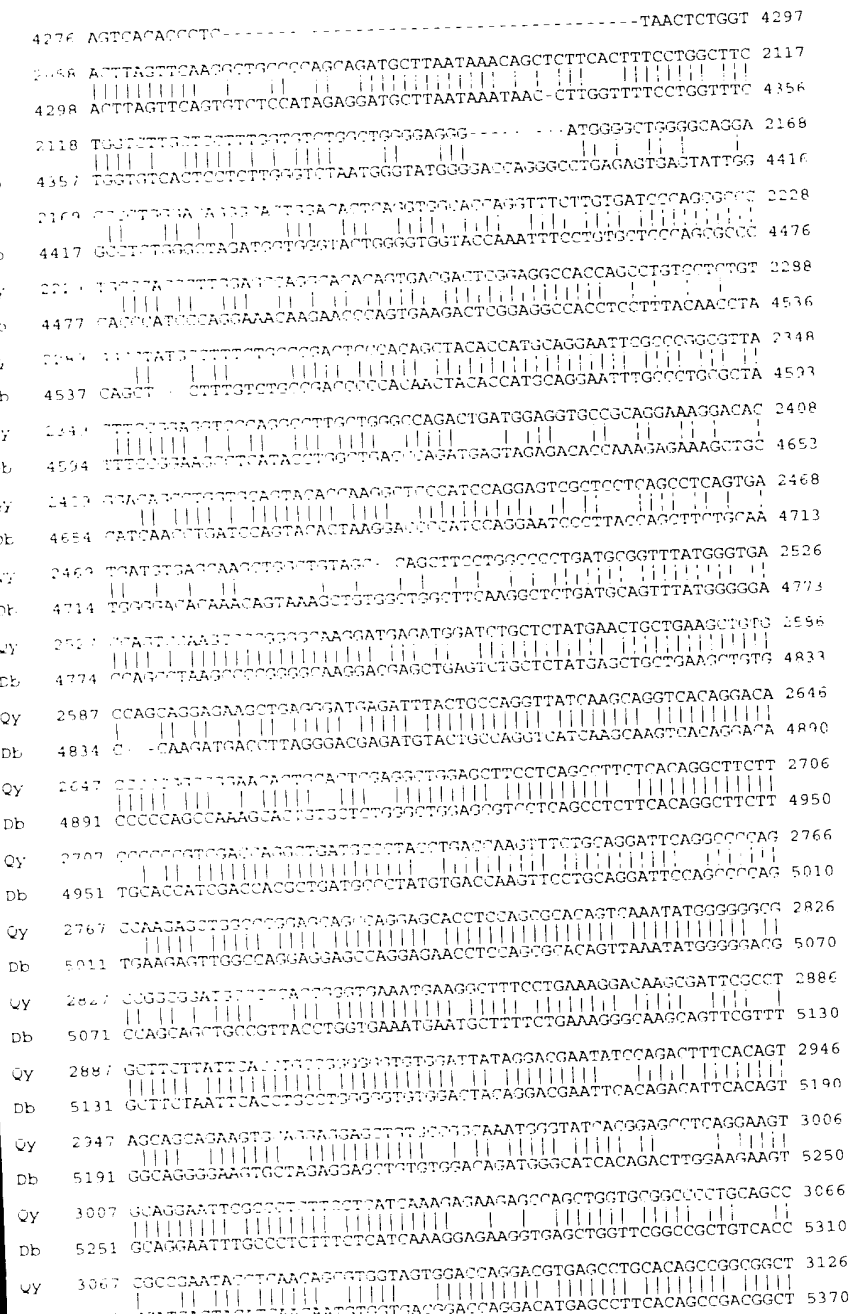
Qy   2298 CTTTCTGCCCGACTCCCACAGCTACACCATGCAGGAATTCCTCCTGTTGGAT 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTTTCTGCCCGACTCCCACAGCTACACCATGCAGGAATTCCTCCTGTTGGAT 50
```

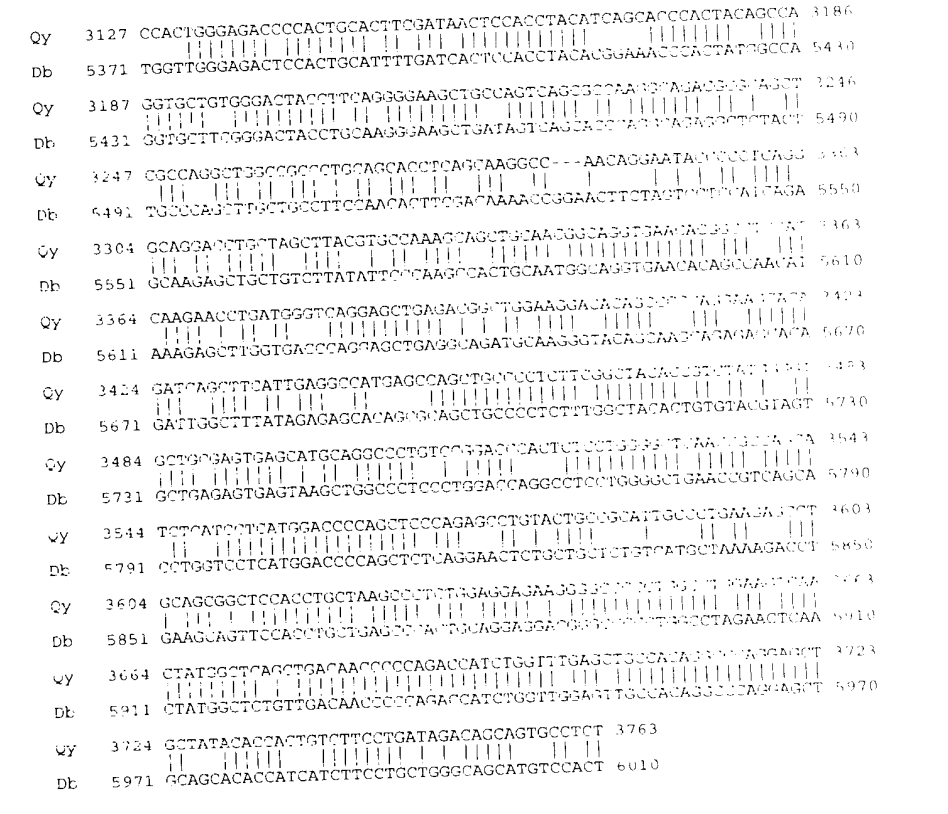




```
RESULT 9
AX253079                                6293 bp    DNA     linear   PAT 05-OCT-2001
LOCUS       AX253079
DEFINITION  Sequence 2 from Patent WO0168866.
ACCESSION   AX253079
VERSION     AX253079.1  GI:15986277
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6293)
  AUTHORS   Brooks,A.F., Deng,G.S. and Rubanyi,G.M.
  TITLE     Estrogen-regulated unconventional myosin-related protein:
            compositions and methods of use
  JOURNAL   Patent: WO 0168866 A 2 20-SEP-2001;
            SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..6293
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
```


/note "Myosin related protein (MPP) variant 1 cDNA"
BASE COUNT 1433 a 1917 c 1733 g 1210 t
ORIGIN
Query Match 42.8%; Score 1787.6; DB 6; Length 6293;
Best Local Similarity 72.3%; Pred. No. 0;
Matches 2685; Conservative 0; Mismatches 969; Indels 166; Gaps 15;





```
RESULT 10
HSM803107
LOCUS       HSM803107                1767 bp    mRNA    linear   PRI 20-MAR-2002
DEFINITION  Homo sapiens mRNA; cDNA DKFZp667C0210 (from clone DKFZp667C0210).
ACCESSION   AL713730
VERSION     AL713730.1  GI:19584466
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia, Eutheria, Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1767)
  AUTHORS   Duesterhoeft,A , Lauber,J., Mewes,H.W., Weil,B. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp667C0210) is available at the RZPD in Berlin.
            Please contact the RZPD: Pessourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY, Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .1767
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="DKFZp667C0210"
                     /tissue_type="lymph node"
                     /clone_lib="667 (synonym: hlno2) Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     gene            1..>1767
                     /gene="DKFZp667C0210"
     CDS             <1..>1386
                     /gene="DKFZp667C0210"
                     /note="similarity to unconventional myosin-15 (Homo
                     sapiens)"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD28520.1"
                     /db_xref="GI:19584467"
                     /translation="DTDSLVQYTKAPIQESLLSLSDDVSKLAVASFLALMEFMGDQSK
                     PRGKDEMDLLYELLKLCQQEELPDEIYQLVIPQVTGHPPPPPHCTPGWSPLSLLTGPPP
                     PSTRLMPYLTKFLQLSGPSQELARSSQEHLQRTVKYGGPPRPMPPPGEMKAFLKGQAIR
                     LLLIHLPGSVDYPTNICTFTVAAFVQPRLCTQMDITEPQEVQFFAIPLIPPPSPLVSP
                     LQPAEYLNSVVVDQDVSLHSPRLHWETPLHFDNSTYISTHYSQVLWDYLQGKIPVSAK
                     ADALLARLAALPHLSPANPNTPQSQLLAYVFPQLPLVNTASIPDLMSQELFLEGTH
                     SPQEAQISFIEAMSQLPLFGYTVYGVLRVSMQALSGPTLLGLNKQHLILMDPSSQSLY
                     CRIALFSTQPHLHLLSFIEEPGPPDIEVNYRSAINPLTIWFSIPLALRILYITVPLIS
                     SASCTEWPSIN"
     polyA_signal    1731..1736
                     /gene="DKFZp667C0210"
     polyA_site      1747
                     /gene="DKFZp667C0210"
BASE COUNT      118 a    559 c    703 g    287 t
ORIGIN
```

Query Match 42.0%; Score 1751.2; DB 9; Length 1767;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2764; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy    2401 GAACATCGACAGCCTGGTGCAGTACACCAAGGCTGCCATCCAGGAGTCGCTCCTCAGCCTC 2460
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GATACTGACAGCCTGGTGCAGTACACCAAGGCTGCCATCCAGGAGTCGCTCCTCAGCCTC 60

Qy    2461 ACTGATGATGTGAGCAAGCTGGCTGTAGCCAGCTTCCTGGCCCCTGATGCGGTTTATGGG 2520
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      61 ACTGATGATGTGAGCAAGCTGGCTGTAGCCAGCTTCCTGG-CCCTGATGCGGTTTATGGG 119

Qy    2521 TCAACAGTCCAAGCCCCGGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAAGCT 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 TCAACAGTCCAAGCCCCGGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAAGCT 179

Qy    2581 GTTCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGGTCACAGG 2640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 GTTCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGGTCACAGG 239

Qy    2641 ACAGCCCGGCCGGAACACTGCACTCGAGGCTGGAGCTTCCTCAGCCTTCTCACAGGCTT 2700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 ACAGCCCGGCCGGAACACTGCACTCGAGGCTGGAGCTTCCTCAGCCTTCTCACAGGCTT 299

Qy    2701 CTTCTCCGCCGTGGACCAGGCTGATGCCCTACCTGACCAAGTTTCTGCAGGATTCAGGCCC 2760
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 CTTCTCCGCCGTGGACCAGGCTGATGCCCTACCTGACCAAGTTTCTGCAGGATTCAGGCCC 359

Qy    2761 AAGTAAAAGTGTGGCCTGGAGCAGCTAGGAGCACTTCCAGCGCACAGTCAAATATGGGGG 2820
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 CAGCCAAGACGTGGCCCTGGAGCAGCCAGGAGCACTTCCAGCGCACAGTCAAATATGGGGG 419

Qy    2821 GCGCCGAGCGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGATTCG 2880
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GCGCCGAGCGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGATTCG 479

Qy    2881 CCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGAATATCCAGACTTTCAC 2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 CCTGCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATATCCAGACTTTCAC 539
```

```
Qy    2941 AGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCGGGCAAATGCTTACTACGAGCGCTACGA 3000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 AGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCGGGCAAATGCTTACTACGAGCGCTACGA 599

Qy    3001 AGTGCAGGAATTCGCCCTTCTTCCTCATCAAAGAGAAGAGCCAGTGTGTGGCGCCTGCA 3060
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 AGTGCAGGAATTCGCCCTTCTTCCTCATCAAAGAGAAGAGCCAGTGTGTGGCGCCTGCA 659

Qy    3061 GCCCGCGGAATACCTTAACAGCGTGCTACTGGACCAGGACGCCAGCTTCACAGCCGAGC 3120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 GCCCGCGGAATACCTTAACAGCGTGCTACTGGACCAGGACGCCAGCTTCACAGCCGAGC 719

Qy    3121 GCTCCACTTGGAGACCCCACTGCACTTGGATAACTCCACCTACATCAGCACCCACTACAG 3180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 GCTCCACTTGGAGACCCCACTGCACTTGGATAACTCCACCTACATCAGCACCCACTACAG 779

Qy    3181 CCAGGTGCTGTGGGACTACCTGCAGGGGAAGCTGCCAGTGAGCGCCAAGGCGGACGCCT 3240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 CCAGGTGCTGTGGGACTACCTGCAGGGGAAGCTGCCAGTGAGCGCCAAGGCGGACGCCT 839

Qy    3241 GCTGCTGGCAGGCTGGGCGCTGCTGCTGGCCAAGCTCAGCAAGGCCGAAGAGGAATACC 3300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 GCTGCTGGCAGGCTGGGCGCTGCTGCTGGCCAAGCTCAGCAAGGCCGAAGAGGAATACC 899

Qy    3301 GCGCCGAGGTGGGCAAGGTGTGCGCCCTGGAGGAGCTGACCCAGTACCTGGAGGAGCTGC 3360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GCAGGAGCTGCTAGCTTACGTGCCAAAGCAGCTGCAAGCGGTGCTGACAGCCTCTCTAT 959

Qy    3361 CAAGAACCTGATGGGTCAGGAGCTGACGGCCTGGAAGGAGCTACAGCCCCAGGAGCTGCA 3420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 CAAGAACCTGATGGGTCAGGAGCTGACGGCCTGGAAGGAGCTACAGCCCCAGGAGCTGCA 1019

Qy    3421 GATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCCCTATCGGCT 3480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 GATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCCCTATCGGCT 1079

Qy    3481 GCTGCGAGTGAGGCATGCAGGCCCTGTCCGGACCCACTCTCCCCAACCCCAGCCCA 3540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 GCTGCGAGTGAGGCATGCAGGCCCTGTCCGGACCCACTCTCCCCAACCCCAGCCCA 1139

Qy    3541 TCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACCTGCCCATCCCCAAGAAGCCT 3600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 TCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACCTGCCCATCCCCAAGAAGCCT 1199

Qy    3601 GCAGCGGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGCCCTGAAGCCTAA 3660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 GCAGCGGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGCCCTGAAGCCTAA 1259

Qy    3661 CTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGACGTGCTAAGCGGCAAGGACGT 3720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 CTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGACGTGCTAAGCGGCAAGGACGT 1319

Qy    3721 GCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCATCAGCGTGCCCAGCAT 3780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 GCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCATCAGCGTGCCCAGCAT 1379

Qy    3781 CAACTGAGAGGAGTGCAGGCCGGGGAGACGAAGAGGATGAGGCTCCCCGGCCCAAGTCT 3840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 CAACTGAGAGGAGTGCAGGCCGGGGAGACGAAGAGGATGAGGCTCCCCGGCCCAAGTCT 1439

Qy    3841 CACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTTCTAGAACCTGCTTCAGTACA 3900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 CACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTTCTAGAACCTGCTTCAGTACA 1499

Qy    3901 GCCCAGCCGGGCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGCCACCAGCAGGCA 3960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 GCCCAGCCGGGCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGCCACCAGCAGGCA 1559

Qy    3961 AAGAAAACAGCCAGACCCCTCTCCAGCAGGCCTGGGGGCCAAAGCCGGGTCTGCAGAAACTCG 4020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 AAGAAAACAGCCAGACCCCTCTCCAGCAGGCCTGGGGGCCAAAGCCGGGTCTGCAGAAACTCG 1619
```

```
Qy   4024 GCTGGGGCACCTGAGGTTGGGCAGTCTGAGGGAGATGCCCACCCGACCCCAGGCTCCGCC 4083
Db   1620 GCTGGGGCACCTGAGGTTGGGCAGTCTGAGGGAGATGCCCACCCGACCCCAGGCTCCGCC 1679
Qy   4084 CAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAACT 4143
Db   1680 CAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAACT 1739
Qy   4144 CCCTAAAAAAAAAAAAAAAAAAAAAAAAA 4171
Db   1740 CCCTGAGAAAAAAAAAAAAAAAAAAAAAA 1767

RESULT 11
AF418290
LOCUS       AF418290                 7925 bp    mRNA    linear   PRI 06-JUN-2002
DEFINITION  Homo sapiens clone 15 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418290
VERSION     AF418290.1  GI:16303813
KEYWORDS    .
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7925)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21960014
  PUBMED    11964073
REFERENCE   2  (bases 1 to 7925)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..7925
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="15"
                     /tissue_type="stomach"
     gene            <1..>7925
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT     1454 a   2543 c   2691 g   1237 t
ORIGIN

  Query Match          40.5%;  Score 1691.4;  DB 9;  Length 7925;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1714;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy      1 GGGCAGCAGCAGGCTCGGGCTTCGCAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACC 60
Db   6210 GGGCAGCAGCAGGCTCGGGCTTCGCAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACC 6269
Qy     61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120
Db   6270 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 6329
Qy    121 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
Db   6330 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 6389
Qy    181 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
Db   6390 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 6449
Qy    241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG 300

Db   6450 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG 6509
Qy    301 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGGCGTCCCCT 360
Db   6510 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGGCGTCCCCT 6569
Qy    361 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
Db   6570 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 6629
Qy    421 GCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480
Db   6630 GCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 6689
Qy    481 GGCAGGGGGACTGTGGTGCGCAG-TCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGG 539
Db   6690 GGCAGGGGGACTGTGGTGCGCAGCTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGG 6749
Qy    540 GAAATTGGCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCCGTG 599
Db   6750 GAAATTGGCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCCGTG 6809
Qy    600 CAGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGCCTCTG 659
Db   6810 CAGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGCCTCTG 6869
Qy    660 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTGTGCCCCAGCCCAGGA 719
Db   6870 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTGTGCCCCAGCCCAGGA 6929
Qy    720 AAGGGCCCCCCGCCAGCTGTGGCTCCTCGACCCAAGGCCCCGCTACAGCTTGGGCCCTCT 779
Db   6930 AAGGGCCCCCCGCCAGCTGTGGCTCCTCGACCCAAGGCCCCGCTACAGCTTGGGCCCTCT 6989
Qy    780 AGCTCCATCAAGGAAAAGCAGGGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATT 839
Db   6990 AGCTCCATCAAGGAAAAGCAGGGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATT 7049
Qy    840 GCCCACACACCCCCACCTCCACCAGGGCCACCACTGCCTCTGCCCGAGGACCCAGGGACC 899
Db   7050 GCCCACACACCCCCACCTCCACCAGGGCCACCACTGCCTCTGCCCGAGGACCCAGGGACC 7109
Qy    900 CTTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAG 959
Db   7110 CTTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAG 7169
Qy    960 CTACTCGCGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACGCCCTGGAAGTTGTTC 1019
Db   7170 CTACTCGCGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACGCCCTGGAAGTTGTTC 7229
Qy   1020 CTACGCAAGGAGGTGTTCTACCCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCTC 1079
Db   7230 CTACGCAA-GAGGTGTTCTACCCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCTC 7288
Qy   1080 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTATCCGGATTTCCCAGAAT 1139
Db   7289 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTATCCGGATTTCCCAGAAT 7348
Qy   1140 GAGCGGCGGAAAATGAAAGACCTGCTGGAGGCTTGGAGGTGGACCTGGATTCTCTCAGC 1199
Db   7349 GAGCGGCGGAAAATGAAAGACCTGCTGGAGGCTTGGAGGTGGACCTGGATTCTCTCAGC 7408
Qy   1200 ACCACCGAAGACAGCGTCAAGAAGGGCATCGTGGTGTCCGGGGGAGAGTGGCAGCGACCT 1259
Db   7409 ACCACCGAAGACAGCGTCAAGAAGGGCATCGTGGTGTCCGGGGGAGAGTGGCAGCGACCT 7468
Qy   1260 TACTTCTCCCGCTTCTTTCCTGTCTCGGGGGAGAGTGGCAGCGACCTGCAGCTGTTAGCC 1319
Db   7469 TACTTCTCCCGCTTCTTTCCTGTCTCGGGGGAGAGTGGCAGCGACCTGCAGCTGTTAGCC 7528
Qy   1320 GTGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGGCCCCCCCCCCCCCCCCGAC 1379
```

```
Db    7529 ATATCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGGCCCCGGCCTCCGCCCCGAC 7588

Qy    1380 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGTGTGAGTGCCGGGGC 1439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7589 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGTGGAGTGCCGGGGC 7648

Qy    1440 AGTCCAGCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTGCACACAGCCGGGGCA 1499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7649 AGTCCAGCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTGCACACAGCCGGGGCA 7708

Qy    1500 AGGGCATCGAGGGGCTGGTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTAT 1559
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7709 AGGGCATCGAGGGGCTGGTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTAT 7768

Qy    1560 CTCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCAGCTTCCACCTGGG 1619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7769 CTCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCAGCTTCCACCTGGG 7828

Qy    1620 GAGCTCATCAAGCTGCTGCCGGTGGCTACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCC 1679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7829 GACCTCATCAAGCTGCTGCCGGTGGCTACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCC 7888

Qy    1680 GTGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTGC 1716
           ||||||||||||||||||||||||||||||||||||
Db    7889 GGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTGC 7925


RESULT 12
AK096869
LOCUS       AK096869                1998 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ39550 fis, clone RECTM1000141.
ACCESSION   AK096869
VERSION     AK096869.1  GI:21756455
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens rectum cDNA to mRNA, clone_lib:RECTM1
            clone:RECTM1000141.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nimomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Fukkawa,F., Omura,Y., Abe,F.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Ninoyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Fukuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1998)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan. cDNA full insert sequencing:
            Research Association for Biotechnology; FARE, cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: FARE,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..1998
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RECTM1000141"
                     /tissue_type="rectum"
                     /clone_lib="RECTM1"
                     /note="cloning vector: pME18SFL3"

BASE COUNT      437 a    642 c    596 g    423 t
ORIGIN

  Query Match             88.2%;  Score 1386.6;  DB 4;  Length 1998;
  Best Local Similarity   99.7%;  Pred. No. 4e-266;
  Matches 1389;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    2763 CCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCACCTCCAGGGCAAGCCCAAATATGGGG 2822
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     604 CCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCACCTCCAGGGCAAGCCCAAATATGGGG 663

Qy    2823 GGCGCCGGCGGATGCCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGCAATAACCATTC 2882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     664 GGCGCCGGCGGATGCCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGCAATAACCATTC 723

Qy    2883 GCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGCATAATA  AGCATTTTA 2942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 GCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGCATAATATCAGCATTTTA 783

Qy    2943 CAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTATCAGGAGCTTCAGT 3002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 CAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTATCAGGAGCTTCAGT 843

Qy    3003 AAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCAGCAGCTGGTGGCCCTGC 3062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     844 AAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCAGCAGCTGGTGGCCCTGC 903

Qy    3063 AGCCCGCCGAATACCTCTAACAGCGTGCTAGTCGACGCAGAGACCGAGCCTTCCACAGCGG 3122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     904 AGCCCGCCGAATACCTCTAACAGCGTGCTAGTCGACGCAGAGACCGAGCCTTCCACAGCGG 963

Qy    3123 GGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTATCA 3168
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     964 GGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTATCA 1009

Qy    3169 GCCAGGTGCTGTGGGACTACCTCAGGGAAGCTGCCAGTCAGGCCAAGCCAGACGGC 3224
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 GCCAGGTGCTGTGGGACTACCTCAGGGAAGCTGCCAGTCAGGCCAAGCCAGACGGC 1065

Qy    3225 AGCTCGCCAGGCTGGCCGCCTGCAGCACCTCAGCAAGGCCAACAGGAATAACCCCTCTAG 3284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1066 AGCTCGCCAGGCTGGCCGCCTGCAGCACCTCAGCAAGGCCAACAGGAATAACCCCTCTAG 1125

Qy    3285 GGCAGGACCTGCTAGCTTACGTGCCAAAGTAGCTGCAAGGCAAGTCAGGCTCTCA 3340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1126 GGCAGGACCTGCTAGCTTACGTGCCAAAGTAGCTGCAAGGCAAGTCAGGCTCTCA 1181

Qy    3341 TCAAGAACCTGATGGGTCAGGAGCTGAACGCCTGGAAGGACAACAGCCCCAAGAACCAC 3400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1182 TCAAGAACCTGATGGGTCAGGAGCTGAACGCCTGGAAGGACAACAGCCCCAAGAACCAC 1241

Qy    3401 AGATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTTGGCTTATCAGTGTATGGGG 3460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1242 AGATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTTGGCTTATCAGTGTATGGGG 1301

Qy    3461 TGCTGCGAGTGAGCATGCAGGCCTCTCCGGACCCATCTCCTGAAGCCAAGCAGAGT 3516
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1302 TGCTGCGAGTGAGCATGCAGGCCTCTCCGGACCCATCTCCTGAAGCCAAGCAGAGT 1357

Qy    3517 ATCTTATCTTATGAACCCAGTTCCAGATCCTGTACTTCCATCTCAAAGCC 3566
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1358 ATCTTATCTTATGAACCCAGTTCCAGATCCTGTACTTCCATCTGAAGCC 1407

Qy    3567 TGCAGCGGCTCCACCTGCTAAGCCTCTGGAGAGAAGAAGGGCCCCCTGGAAAGTCA 3622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1408 TGCAGCGGCTCCACCTGCTAAGCCTCTGGAGAGAAGAAGGGCCCCCTGGAAAGTCA 1463

Qy    3623 ACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGACGCTGCTACAGCCCAGCAGC 3682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1464 ACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGACGCTGCTACAGCCCAGCAGC 1523

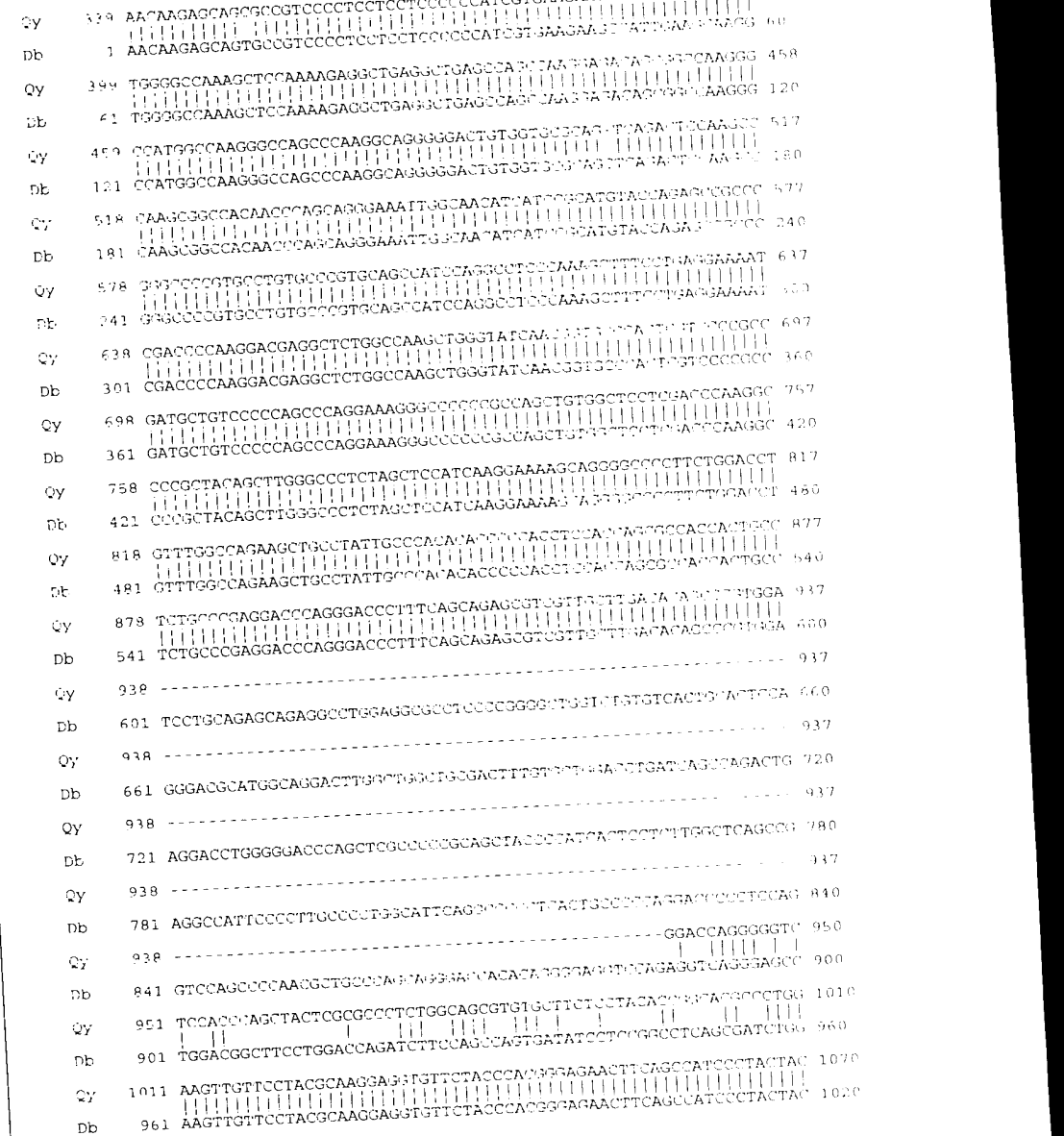
Qy    3683 TGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGACTGAGTGCCTCAGA 3742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1524 TGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGACTGAGTGCCTCAGA 1583
```

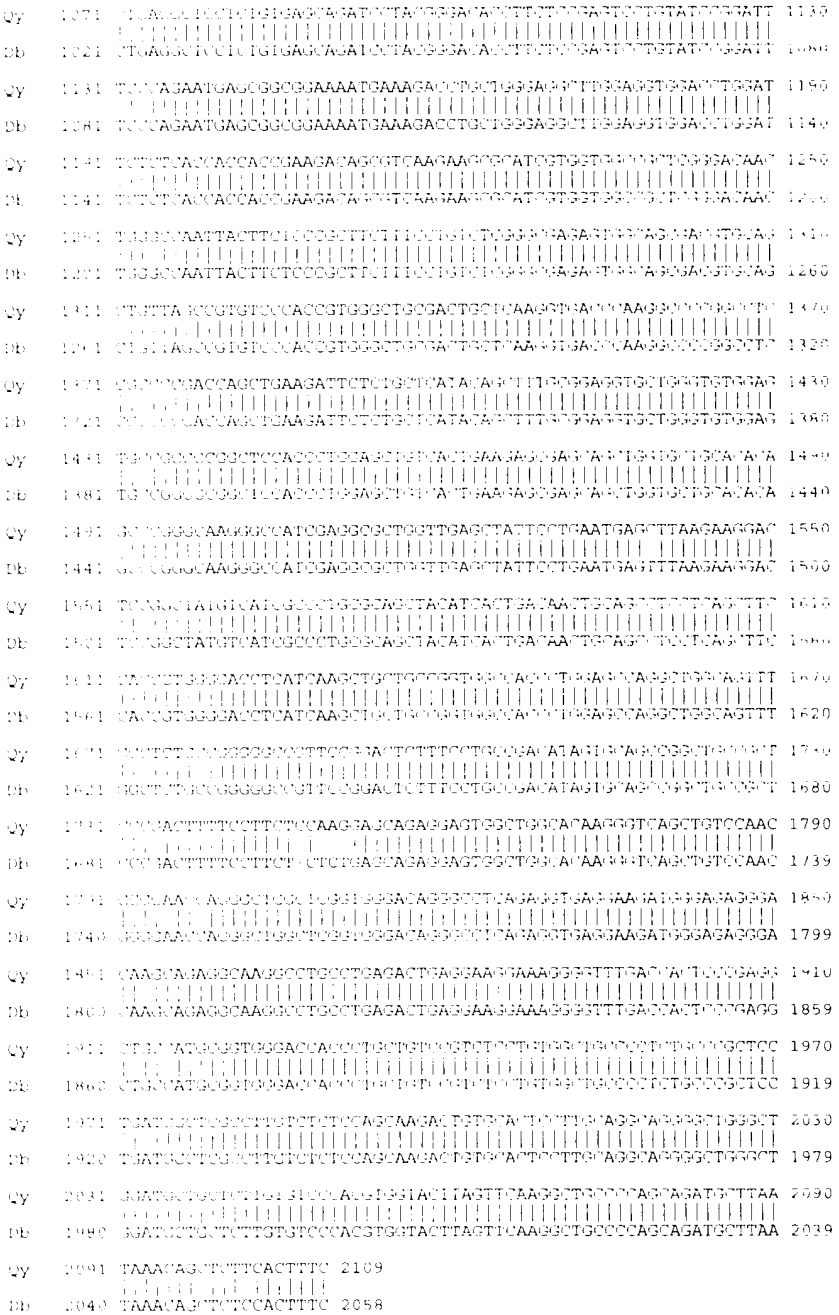


RESULT 13
AK025259 2049 bp mRNA linear PRI 29-SEP-2000
LOCUS AK025259
DEFINITION Homo sapiens cDNA: FLJ21606 fis, clone COL07302.
ACCESSION AK025259
VERSION AK025259.1 GI:10437734
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (clone cDNA to mRNA, clone_lib:COL clone:COL07302.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Wakitani,P., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2049)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108 8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81 3-5449-5286, Fax 81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5' & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency)
FEATURES Location/Qualifiers
source 1..2049
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL07302"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
BASE COUNT 433 a 690 c 606 g 370 t

ORIGIN

Query Match 33.0%; Score 1375.8; DB 9; Length 2099;
Best Local Similarity 83.7%; Pred. No. 5.4e-261;
Matches 1723; Conservative 0; Mismatches 47; Indels 28; Gaps 3;





```
RESULT 14
AF418286
LOCUS       AF418286                7515 bp    mRNA    linear   PRI 06-JUN 2002
DEFINITION  Homo sapiens clone 1 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418286
VERSION     AF418286.1  GI:16303809
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7515)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell Motil. 22 (5), 477-488 (2001)
  MEDLINE   21960014
   PUBMED   11964073
REFERENCE   2  (bases 1 to 7515)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..7515
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="1"
                     /tissue_type="kidney"
     gene            <1..>7515
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT     1350 a   2469 c   2540 g   1156 t
ORIGIN

  Query Match           27.2%;  Score 1135.6;  DB 9;  Length 7515;
  Best Local Similarity   66.0%;  Fred. No. 1.1e-21;
  Matches 1344;  Conservative    0;  Mismatches  124;  Indels   94;  Gaps    4;

Qy     249 CCCGGGCAAAGGTCCACATCCCCCAGGGGGAAGCCCAGAGAGAAGCAGAGAGAAGA 308
           || |||| ||      |  ||||||      | | |   || | || | | | | |
Db    5954 CCTGGCCCAGCAGATGACAGCCCAGCCATGTCCTGTCCCTGAGTAGTAATGCAGTA 6013

Qy     309 GGAGGAGGAGCAGGAGGAGCAAGAAGCCCAAAGAGAGAGCAGCCCCCTCCTCTCT 368
           | | || ||||    | |  ||||||||||    | | | | | | | | | | |
Db    6014 GCGGCAGCAGCAGGCTCGGGCAAGACTGGAAACAAGCCAGGCTCAGTCCTCAGCCTCT 6073

Qy     369 CCCCATCGTGAAGAAGCCATTGA-AGCAAGGTGGGGCCAAAGCTCCAAAAGAGCCTGAGG 427
           | ||||   |||||||    | | |   |  ||   |  | | |  | | | ||
Db    6074 CTCCATGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGAGTGACCTGGGATCAGA 6133

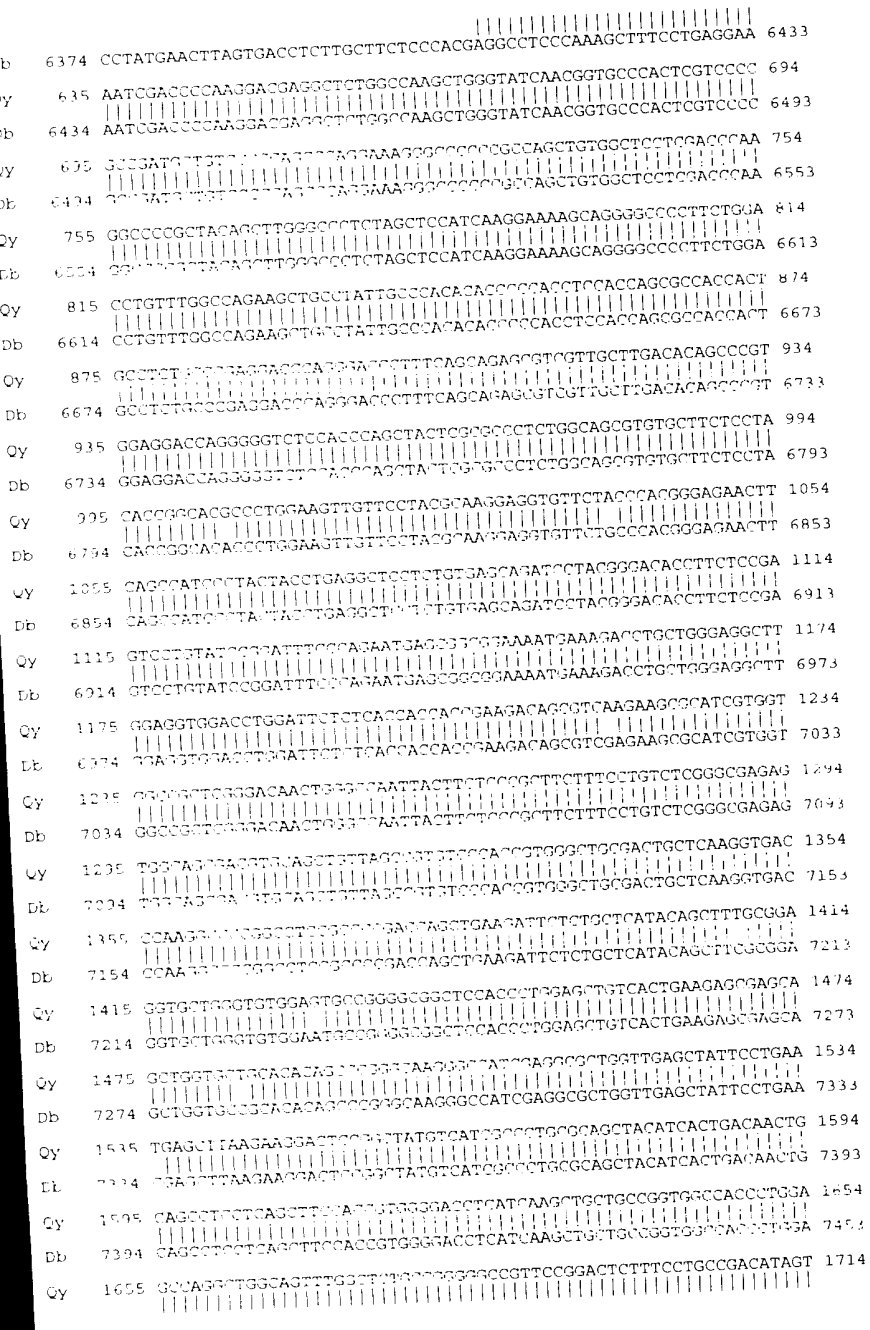
Qy     428 CTGAGCCAGCCAAGGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAAGCCCAAGGC 486
           | | | ||  | || | | | | | | | | |    | | |  | |||||||| |
Db    6134 GGGTGGCTGCCTGAGGGAGACCTCCGAGAGAGCTGAAGACAGGCCCTATAAGCCCAAcc 6193

Qy     487 AGGGGACTGTGGTGCGCAG-TCAGACTCCAAGCCCAAGCCCCCAAAGCCAGCAGGAA 545
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    6194 AGGGGGACTGTGGTGCGCAGCTCAGACTCCAAGCCCAAGCCCCCAAAGCCAGCAGGAA 6253

Qy     546 ATTGGCAAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCTGTGCCTGTGCAGGTGCAG 605
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6254 ATTGGCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCTGTGCCTGTGCAGGTGCAG 6313

Qy     606 CCATCC---------- . . . . . 620
           ||||||
Db    6314 CCATCCAGGTGGGCCCCCAGGGGGGGGGGGCAGGGCTGTCCCTCTCCTGGGGAATTAC 6373

Qy     621 ---------- ---------- -------- --AGGCCTCCCAAAGCTTTCCTGAGGAA 644
```



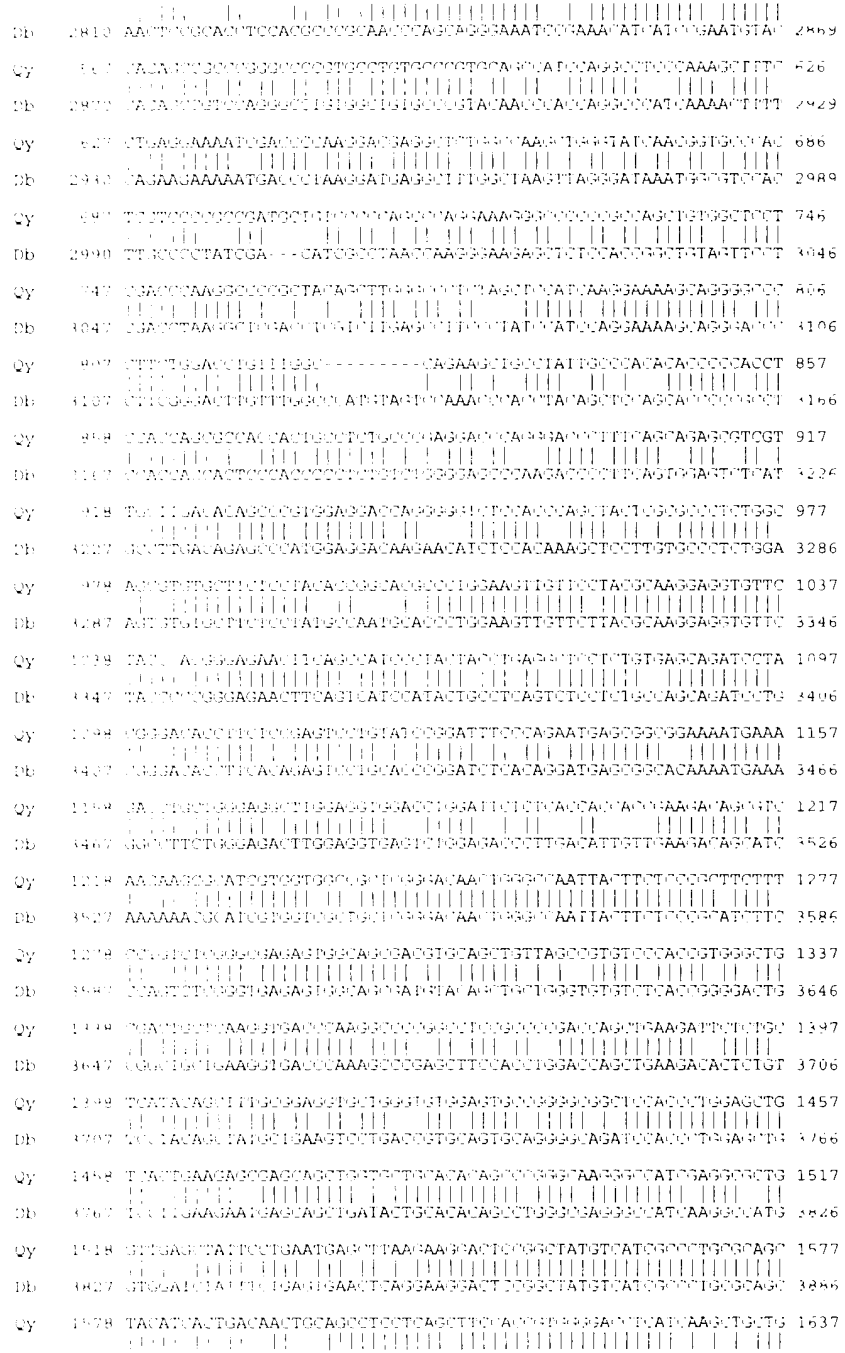
Db 7454 GCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACATAGT 7513

Qy 1715 GC 1716

Db 7514 GC 7515

RESULT 15
AX253080 4375 bp DNA linear PAT 05-OCT-2001
LOCUS AX253080
DEFINITION Sequence 3 from Patent WO0168866
ACCESSION AX253080
VERSION AX253080.1 GI:15986278
KEYWORDS .
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4375)
AUTHORS Brooks,A.R., Deng,G.G. and Rubanyi,G.M.
TITLE Estrogen-regulated unconventional myosin related protein:
compositions and methods of use
JOURNAL Patent: WO 0168866-A 3 20-SEP-2001;
SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1..4375
/organism="Mus sp."
/db_xref="taxon:10095"
/note="Mouse myosin related protein (MRP) variant 2 cDNA"
BASE COUNT 966 a 1368 c 1216 g 825 t
ORIGIN

Query Match 22.7%; Score 946.4; DB 6; Length 4375;
Best Local Similarity 72.9%; Pred. No. 2.5e-176;
Matches 1322; Conservative 0; Mismatches 416; Indels 76; Gaps 5;



Search completed: January 9, 2003, 19:59:23
Job time : 10361.4 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

DM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 02:42:16 ; Search time 726.225 Seconds
(without alignments)
12943.430 Million cell updates/sec

Title: US-09-803-126-5
Perfect score: 4174
Sequence: 1 cgagcagcagcaggctc4433 ... aaaaaaaaaaaaaaaaaaaa 4174

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

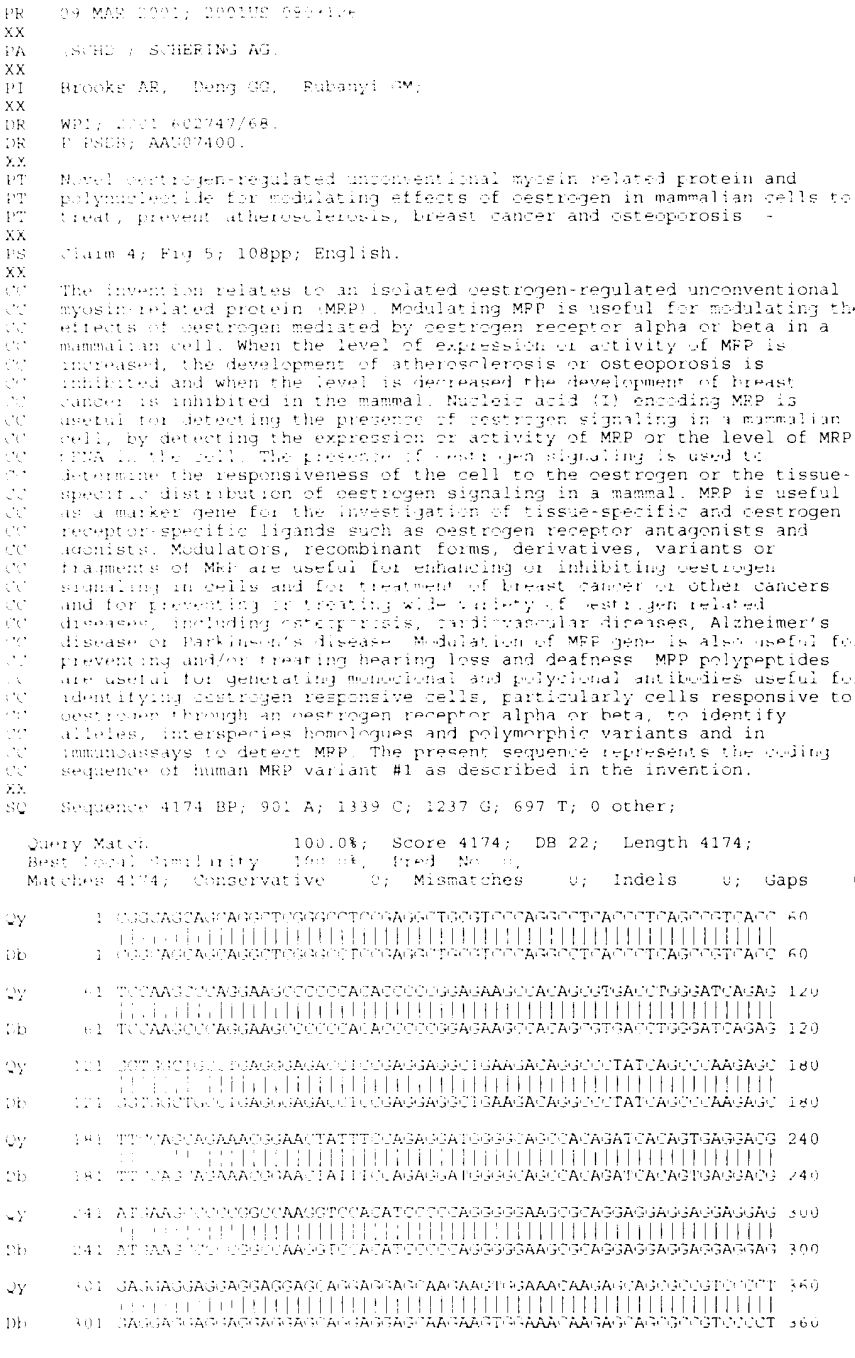
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4174	100.0	4174	22	AAS13821	DNA encoding human
2	3352	80.3	3780	22	AAS13823	DNA encoding human
3	1923.4	46.1	1960	24	ABK35127	Human cDNA encodin
4	1787.6	42.8	6293	22	AAS13819	cDNA #1 encoding m
5	1454	34.8	1476	21	AAZ56716	Human transmembran
6	1241.2	29.7	1268	23	ABK43544	DNA encoding novel
7	1232.2	29.5	1247	22	AAS27110	cDNA encoding nove
8	1232.2	29.5	1247	23	ABK43853	DNA encoding novel
9	1225.8	29.4	1346	22	AAS33274	DNA encoding human
10	946.4	22.7	4375	22	AAS13820	cDNA #2 encoding m
11	715	17.1	24757	22	AAS27687	DNA encoding novel
12	715	17.1	24757	22	AAS33481	DNA encoding human
13	442	10.6	9703	22	AAS33480	Human secreted pro
14	323	7.7	1044	22	AAD07791	Gene #2.6 used to
15	292	7.0	311	24	ABN95022	Colon adenocarcino
16	292	7.0	311	24	ABL62304	Lung cancer relate
17	292	7.0	311	24	ABL64866	Lung cancer relate
18	292	7.0	311	24	ABL65551	Human foetal liver
19	196	4.7	526	22	ABA60490	Probe #7087 for ge
20	196	4.7	526	22	ABA28621	Human brain expres
21	196	4.7	526	22	AAK08667	Human bone marrow
22	196	4.7	526	22	AAK34556	Probe #8365 used t
23	196	4.7	526	22	AAI40274	Human genome deriv
24	196	4.7	526	24	ABS09192	Human foetal liver
25	170	4.1	170	22	ABA72020	Probe #17863 for g
26	170	4.1	170	22	ABA28499	Human brain expres
27	170	4.1	170	22	AAK21359	Human bone marrow
28	170	4.1	170	22	AAK47522	Probe #22039 used
29	170	4.1	170	22	AAI53353	Human genome deriv
30	170	4.1	170	24	ABS21624	Human foetal liver
31	149.8	3.6	233	22	ABA74633	Probe #17863 for g
32	149.8	3.6	233	22	ABA30397	Human brain expres
33	149.8	3.6	233	22	AAK23104	Human bone marrow
34	149.8	3.6	233	22	AAK40279	Probe #23806 used
35	149.8	3.6	233	22	AAI55120	Human genome deriv
36	149.8	3.6	233	24	ABS22813	Human foetal liver
37	149.8	3.6	566	22	ABA62139	Probe #8013 for ge
38	149.8	3.6	566	22	ABA28547	Human brain expres
39	149.8	3.6	566	22	AAK10458	Human bone marrow
40	149.8	3.6	566	22	AAK36359	Probe #10773 used
41	149.8	3.6	566	22	AAI42087	Human genome deriv
42	149.8	3.6	566	24	ABS10413	DNA encoding novel
43	70.4	1.7	509	22	AAS76326	DNA encoding novel
44	61.4	1.5	693	23	AAS74240	DNA encoding novel
45	61.4	1.5	693	23	AAS90715	DNA encoding novel

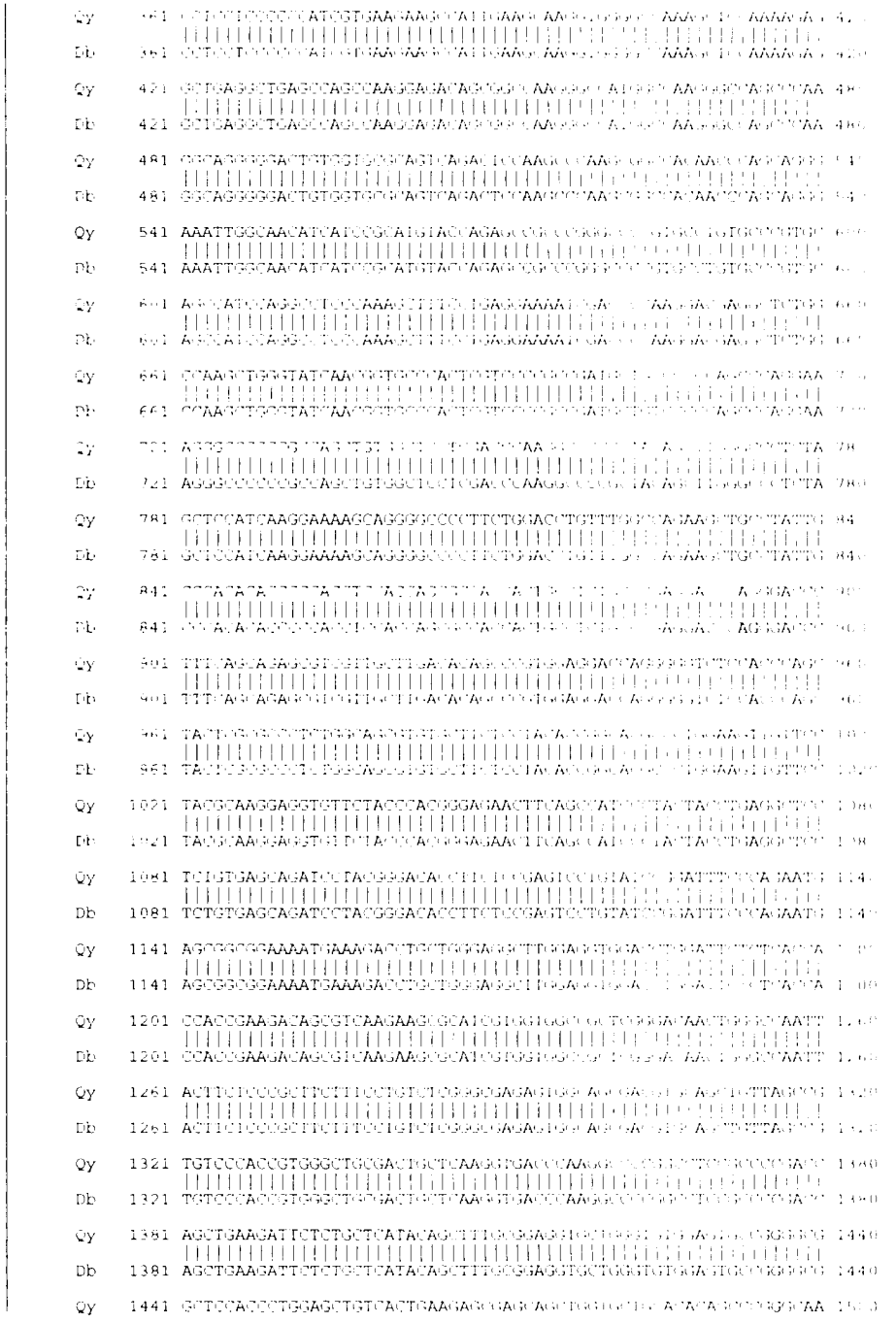
ALIGNMENTS

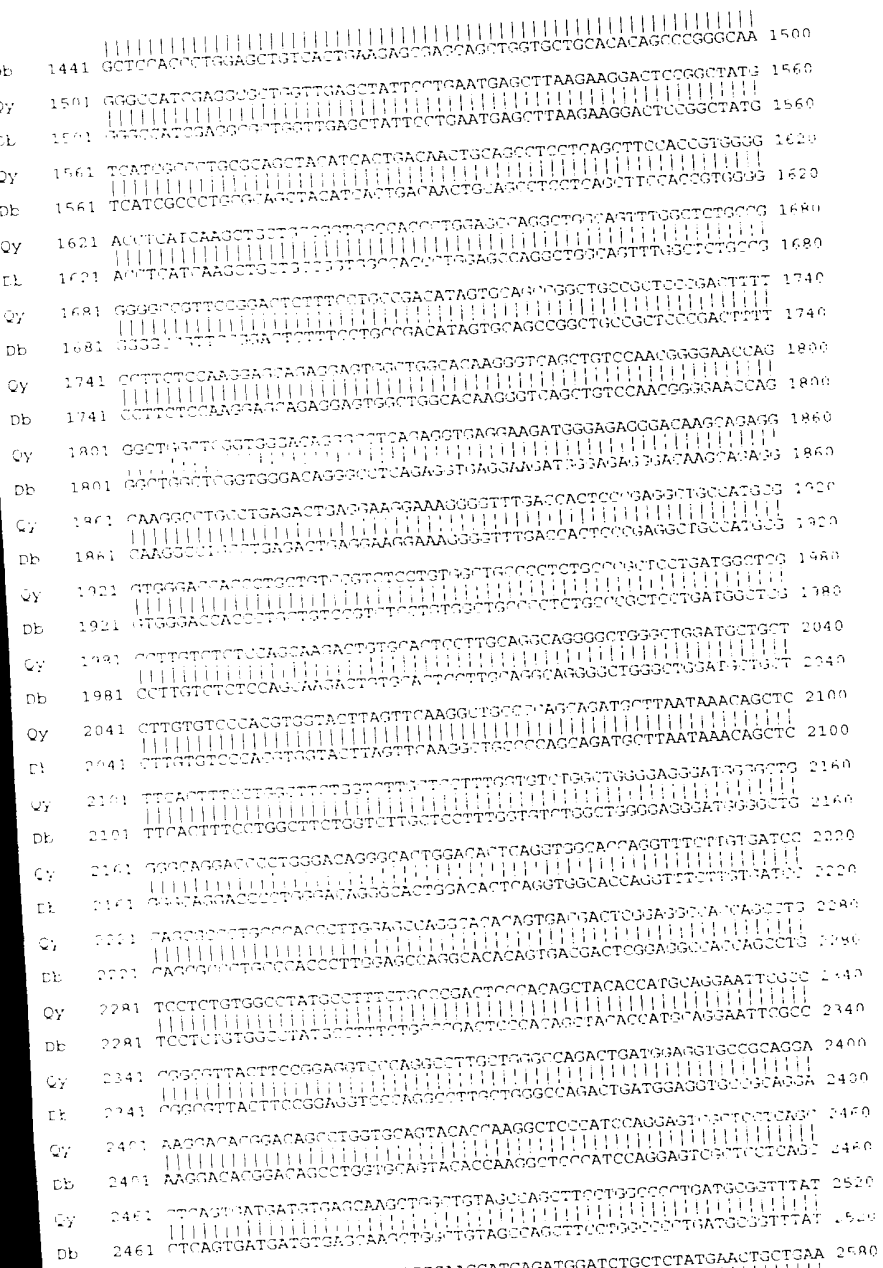
RESULT 1
AAS13821
ID AAS13821 standard; cDNA; 4174 BP.
XX
AC AAS13821;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding human myosin-related protein, hMRP #1.
XX
KW Human; oestrogen-regulated unconventional myosin-related protein; MRP;
KW auditory, antiatherosclerotic; osteopathic, neuroprotective; nootropic;
KW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 561..1874
FT /*tag= a
FT /product= "hMRP"
FT /note= "Oestrogen-regulated unconventional myosin
FT related protein"
FT
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001
XX
PF 12-MAR-2001; 2001WO US08060.
XX
PR 10-MAR-2000; 2000US-188488P.

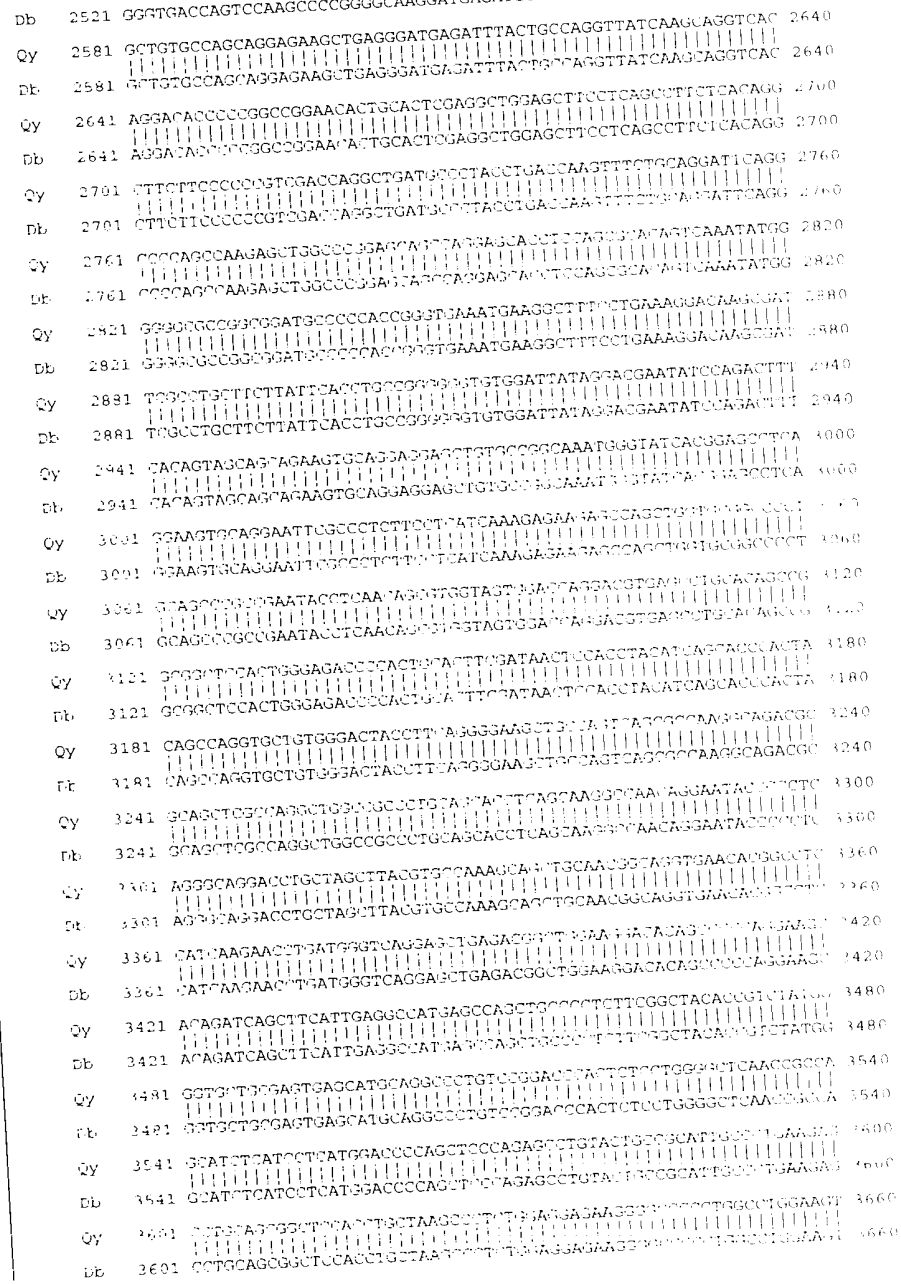
PR 09 MAR 2001; 2001US-09803126.
XX
PA (SCHE) SCHERING AG.
XX
PI Brooks AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR P-PSDB; AAU07400.
XX
PT Novel oestrogen-regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS Claim 4; Fig 5; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating bone loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the coding
CC sequence of human MRP variant #1 as described in the invention.
XX
SQ Sequence 4174 BP; 901 A; 1339 C; 1237 G; 697 T; 0 other;

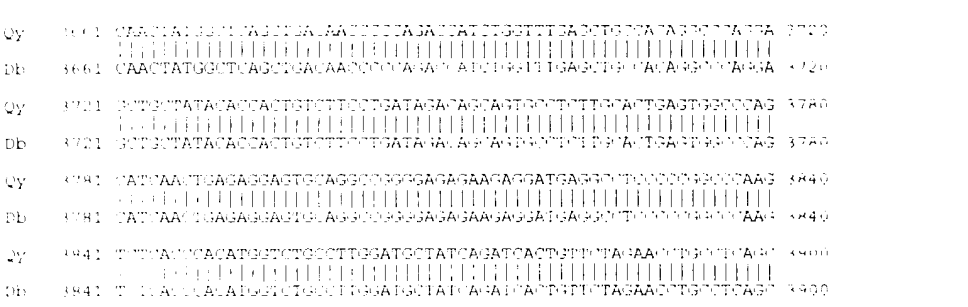
Query Match 100.0%; Score 4174; DB 22; Length 4174;
Best Local Similarity 100.0%; Pred. No. 0.;
Matches 4174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



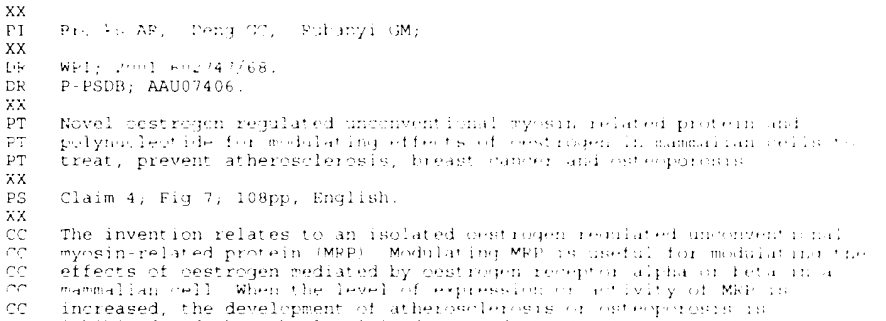


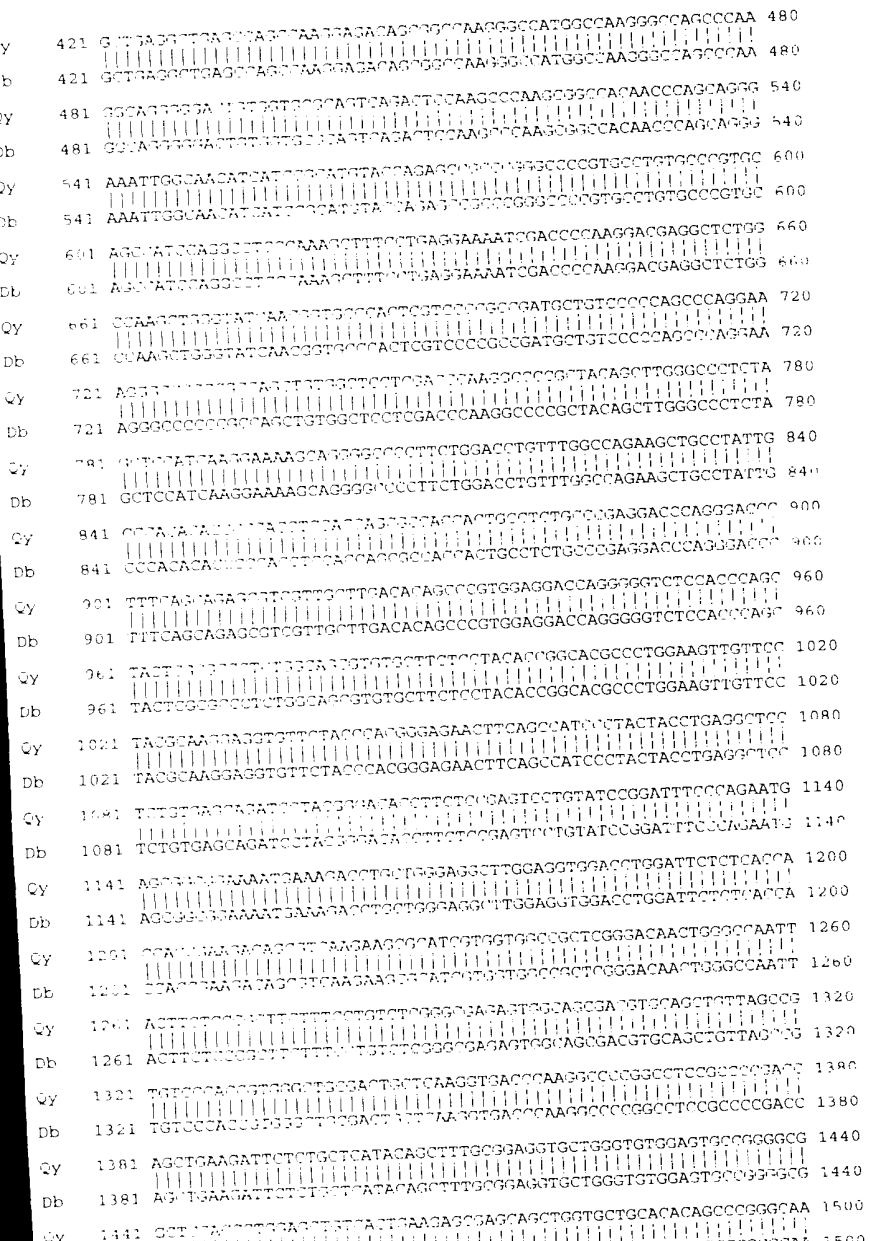


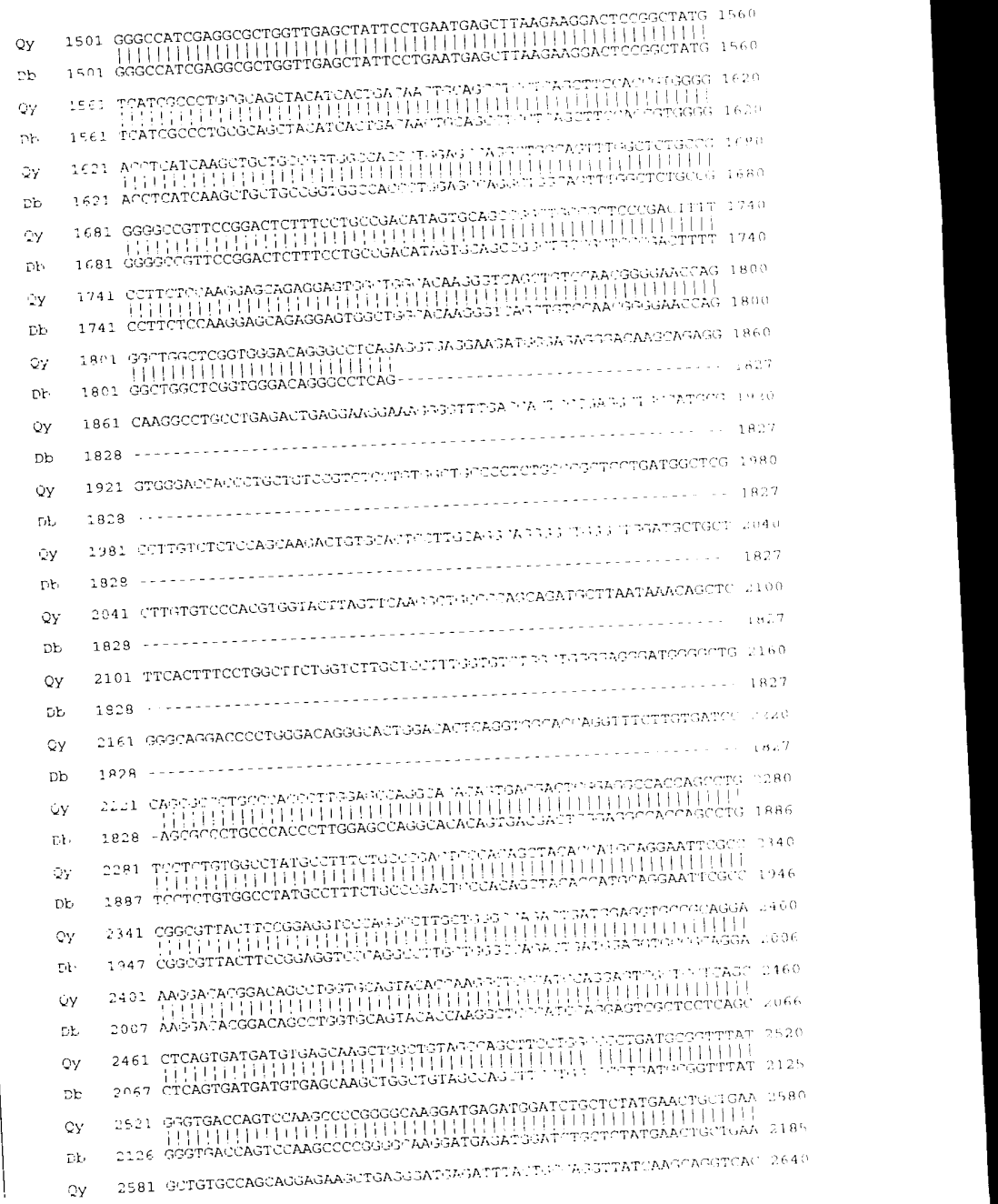


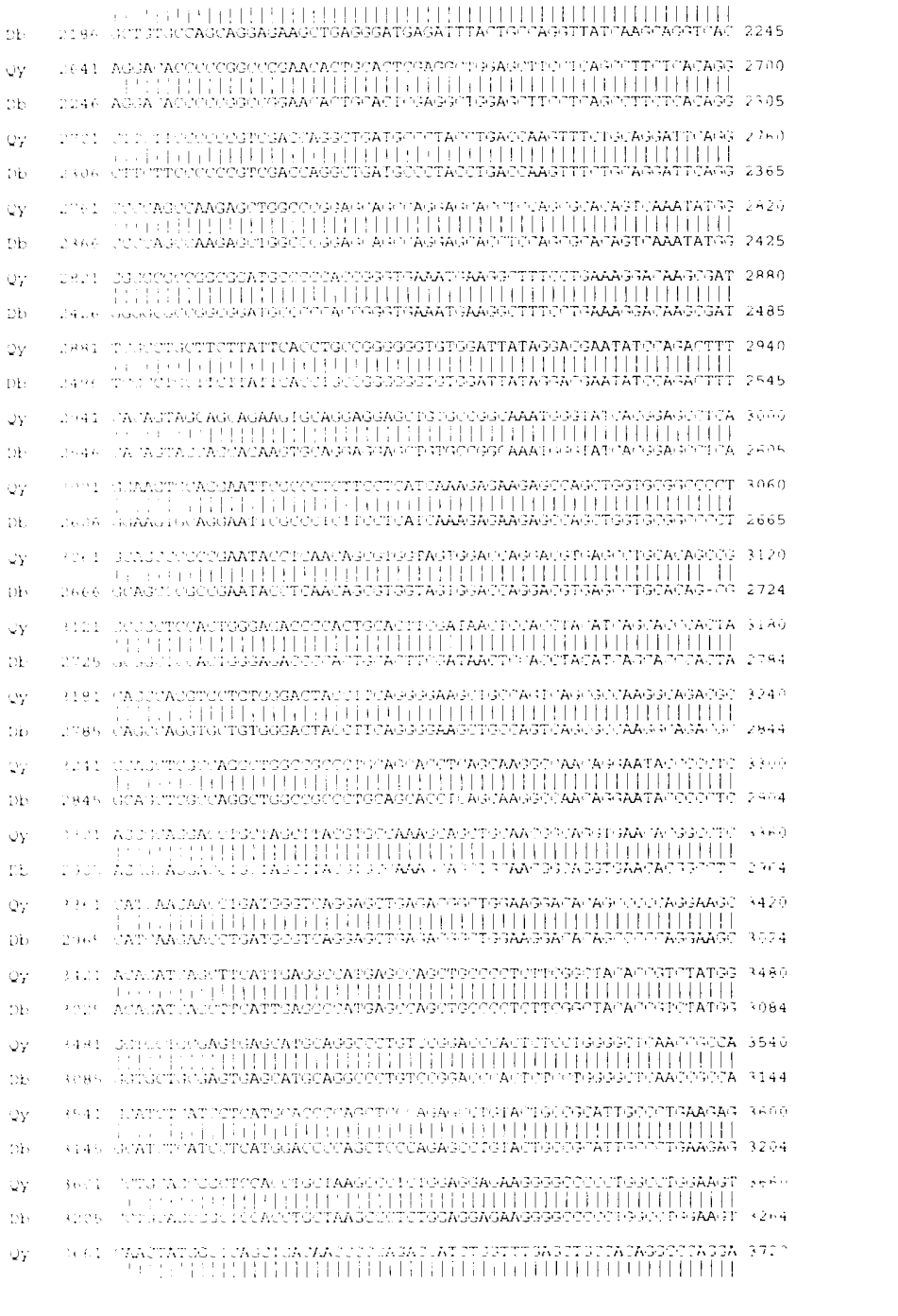


RESULT 2
AAS14823
ID AAS14823 standard; cDNA; 3780 BP.
XX
AC AAS14823;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding human myosin related protein, hMRP #2.
XX
KW Human; oestrogen regulated unconventional myosin-related protein, MRP;
KW inditory; antiatherosclerotic, osteopathic, neurogenerative; nootropic;
KW antiosteoporosis, cytostatic, oestrogen receptor, atherosclerosis,
KW osteoporosis; breast cancer; cardiovascular disease, deafness, ss,
KW Alzheimer's disease, Parkinson's disease, chromosome 17q25, immunogenic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 561..2911
FT /*tag= a
FT /product= "hMRP"
FT /note= "Oestrogen regulated unconventional myosin
FT related protein"
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 16-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0804126.
XX
PA (SCHO-) SCHERING AG.
XX
PI Pro A, AP, Deng CC, Pohanyi GM;
XX
DR WPI; 2001-582743/68.
DR P-PSDB; AAU07406.
XX
PT Novel oestrogen regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 4; Fig 7; 108pp; English.
XX
CC The invention relates to an isolated oestrogen regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the coding
CC sequence of human MRP variant #2 as described in the invention.
XX
SQ Sequence 3780 BP; 838 A; 1228 C; 1107 G; 607 T; 0 other;





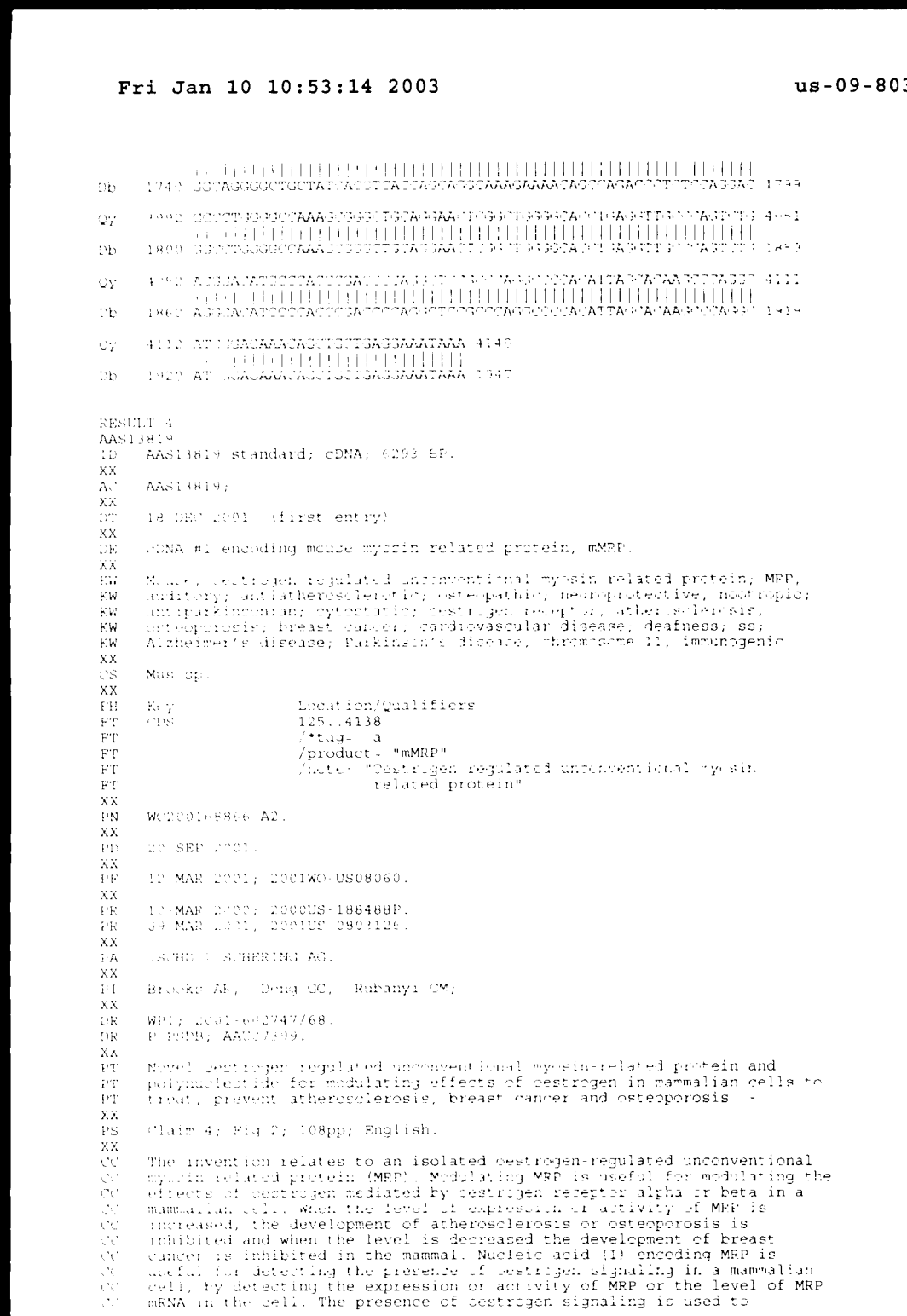




RESULT 3
ABK35127
ID ABK35127 standard; DNA; 1360 BP.
XX
AC ABK35127;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA encoding secreted protein #265.
XX
KW Human; secreted protein; gene; ss; nutritional supplement; hemophilia;
KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW coagulation disorder, inflammatory disorder; "graft" rejection; infection;
KW tissue regeneration, wound healing, burn, haematopoiesis;
KW myeloid cell deficiency; lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN WO200177288-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10224.
XX
PR 06-APR-2000; 2000US-195582P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Bowes SE, Resnick RJ,
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-179321/23.
XX
PT Five hundred and ninety two polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for

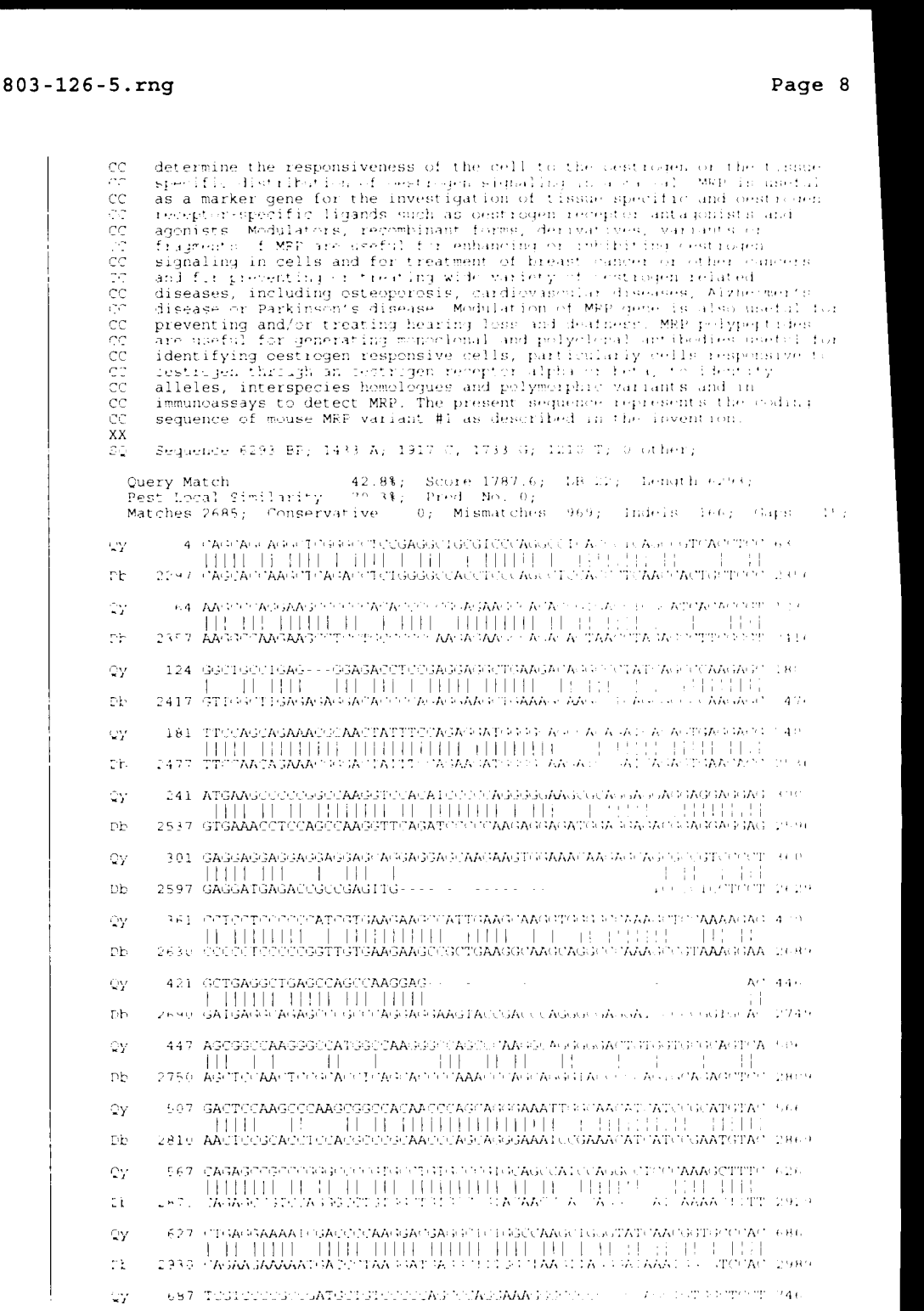
T treating immune deficiencies and disorders such as autoimmune disorders
T -
X
S Claim 1; Page 202-203; 372pp; English.
XX
CC The invention relates to 592 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins. The polynucleotides can be used as probes for the
CC identification and isolation of full length cDNA and genomic DNA. The
CC polynucleotides and proteins can also be used as nutritional supplements.
CC The proteins are useful in the treatment of various immune deficiencies
CC and disorders such as viral infections, bacterial infections, fungal
CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC and conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment of
CC burns, incisions and ulcers. The proteins are also useful for regulating
CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.
XX
SQ Sequence 1960 BP; 433 A; 632 C; 567 G; 328 T; 0 other;

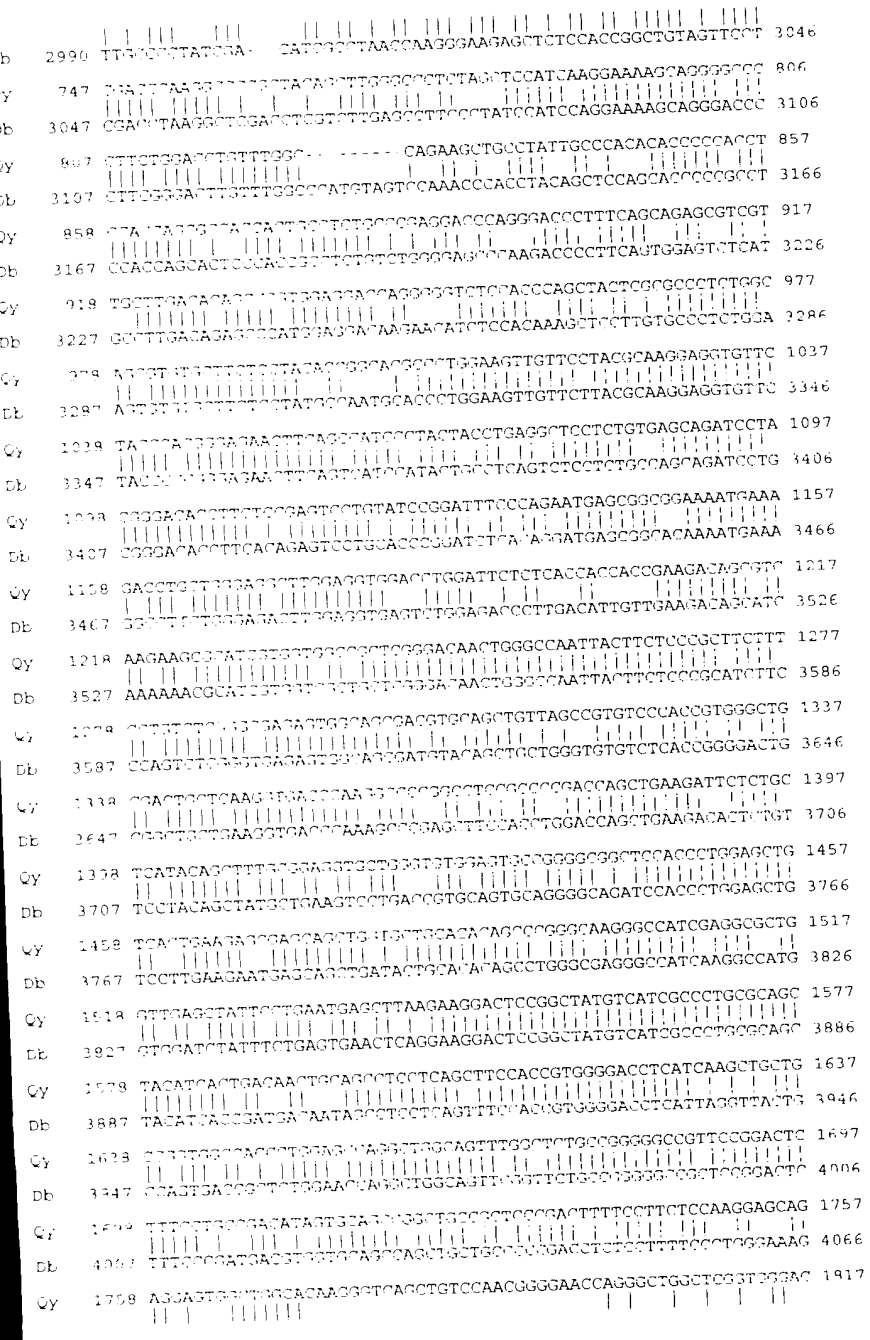
Query Match 46.1%; Score 1923.4; DB 24; Length 1960;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1946; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

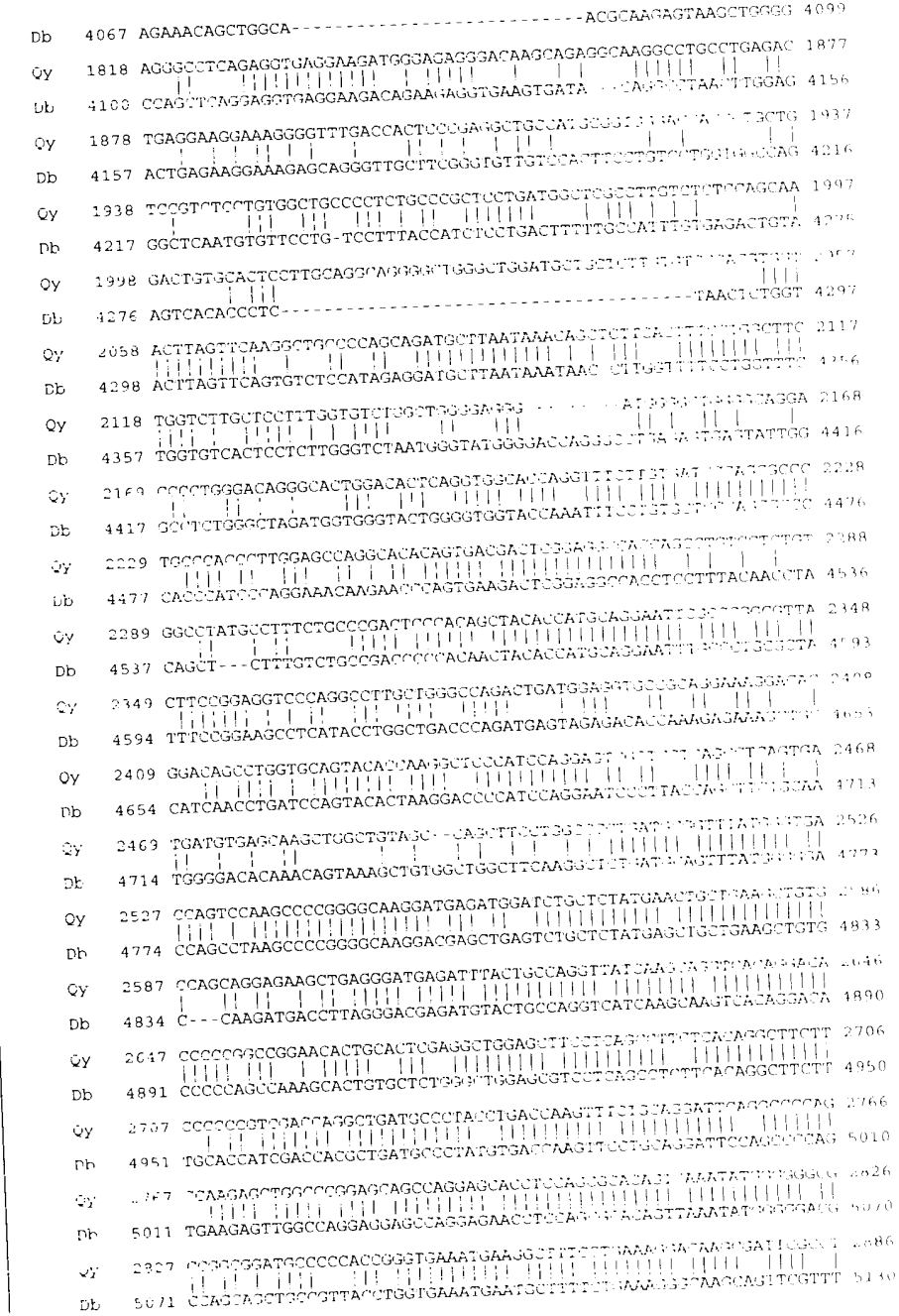


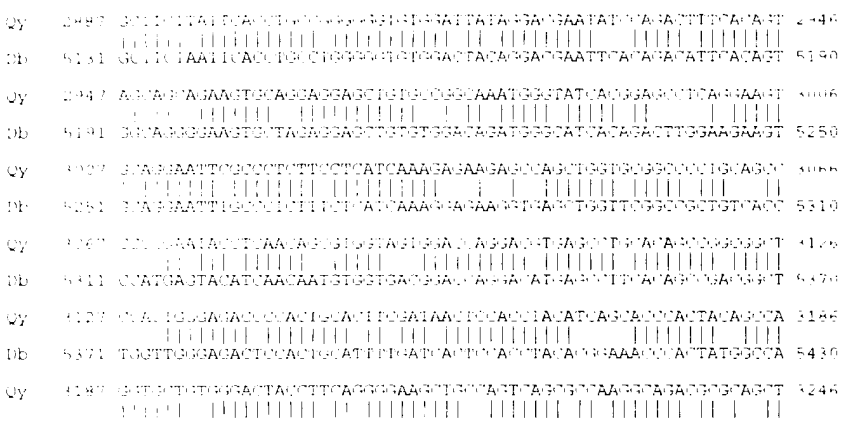
RESULT 4
AAS13819
ID AAS13819 standard; cDNA; 6293 BP.
XX
AC AAS13819;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA #1 encoding mouse myosin related protein, mMRP.
XX
KW Mouse; oestrogen regulated unconventional myosin related protein, MRP;
KW artifact; antiatherosclerotic; osteopathic; neuroprotective; nootropic;
KW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW Alzheimer's disease; Parkinson's disease; chromosome 11; immunogenic.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 125..4138
FT /*tag= a
FT /product= "mMRP"
FT /note= "Oestrogen regulated unconventional myosin
FT related protein"
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 16-MAR-2000; 2000US-188488P.
PR 29-MAR-2001; 2001US-0820126.
XX
PA (SCHE) SCHERING AG.
XX
PI Brooke AS, Dong CC, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR P-PSDB; AAX17499.
XX
PT Novel oestrogen regulated unconventional myosin-related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS Claim 4; Fig 2; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling is also useful. MRP is useful
CC as a marker gene for the investigation of tissue specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating a wide variety of oestrogen related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the coding
CC sequence of mouse MRP variant #1 as described in the invention.
XX
SQ Sequence 6293 BP; 1433 A; 1917 C; 1738 G; 1210 T; 0 other;

Query Match 42.8%; Score 1787.6; DB 22; Length 6293;
Best Local Similarity 70.3%; Pred. No. 0;
Matches 2685; Conservative 0; Mismatches 969; Indels 166; Gaps 17;

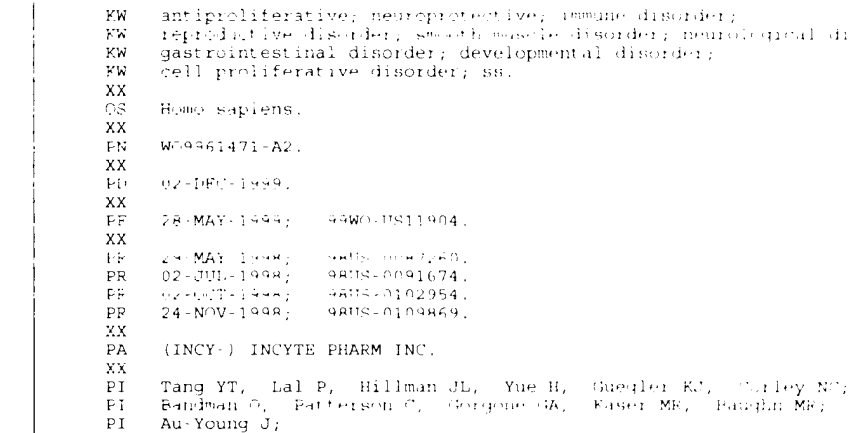






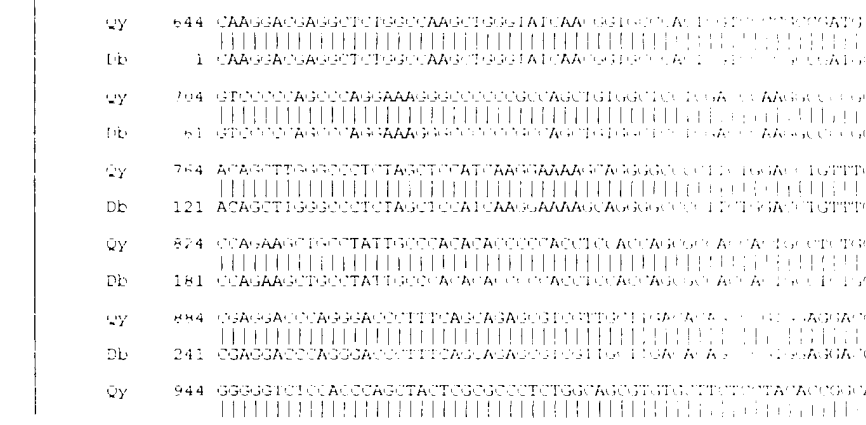


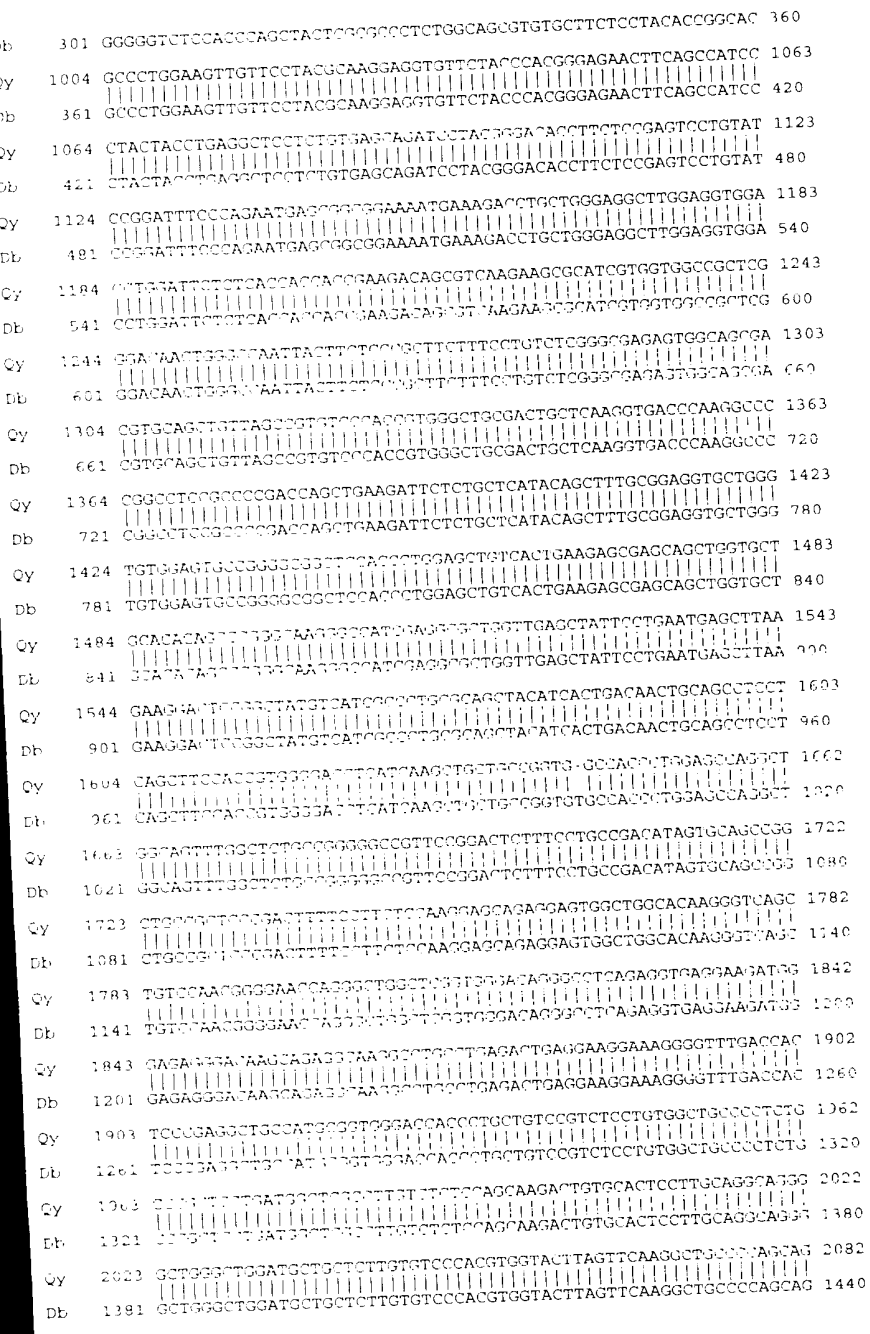
RESULT 5
AAZ56716
ID AAZ56716 standard; cDNA; 1476 BP.
XX
AC AAZ56716;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN 19 encoding cDNA.
XX
KW Human; transmembrane protein; HTMPN; diagnosis; immunospecific;



KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder; ss.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11904.
XX
PR 28-MAY-1998; 98US-0087260.
PR 02-JUL-1998; 98US-0091674.
PR 02-OCT-1998; 98US-0102954.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Gueqler KJ, Corley NC;
PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Ramkumar MR;
PI Au-Young J;
XX
DR WPI; 2000-072605/06.
DR P-PSDB; AAY57895.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders -
XX
PS Claim 9; Page 189; 229pp; English.
XX
CC AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN 1 to HTMPN 79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or
CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.
XX
SQ Sequence 1476 BP; 277 A; 479 C; 435 G; 285 T; 0 other;

Query Match 44.8%; Score 1454; DB 2; Length 1476;
Best Local Similarity 99.9%; Pred. No. 1.1e-281;
Matches 1465; Conservative 0; Mismatches 0; Indels 1; Gaps 1;





Qy 2093 ATGCTTAATAAACAGCTCTTCACTTT 2108
Db 1441 ATGCTTAATAAACAGCTCTTCACTTT 1466

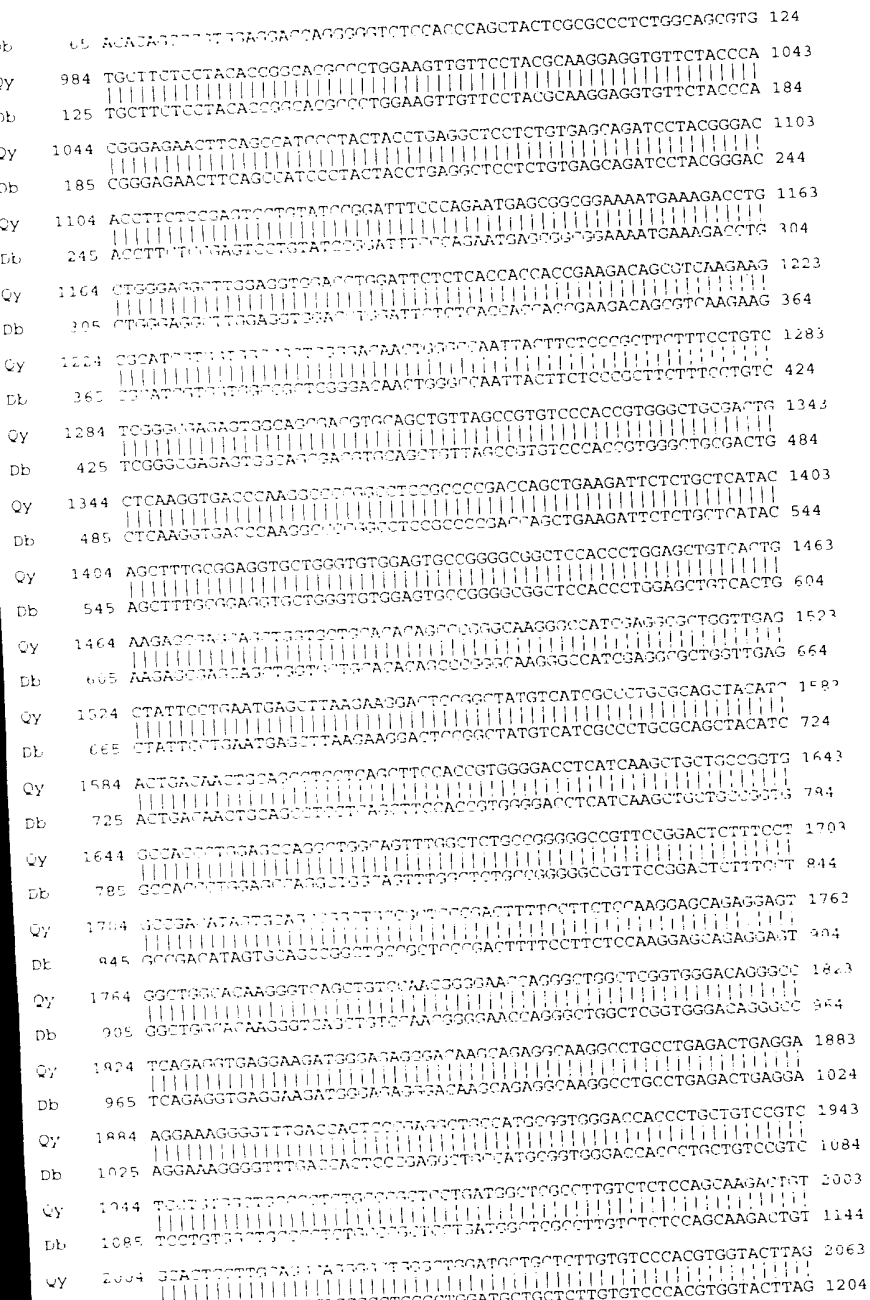
RESULT 6
ABK43544
ID ABK43544 standard; cDNA; 1268 BP.
XX
AC ABK43544;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #124.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PP 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 03-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214986.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229924.

PR 21-SEP-2000; 2000US-0229344.
PR 21-SEP-2000; 2000US-0229345.
PR 25-SEP-2000; 2000US-0229509.
PR 25-SEP-2000; 2000US-0229512.
PR 26-SEP-2000; 2000US-0230437.
PR 26-SEP-2000; 2000US-0230439.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231980.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234273.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234497.
PR 25-SEP-2000; 2000US-0234498.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0236327.
PR 28-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241820.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR P-PSDB; AAU87214.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 1; SEQ ID No 134; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein (II) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast, liver, and vascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyolutrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving immune clarification e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 29.7%; Score 1241.2; DB 23; Length 1208;
Best Local Similarity 99.8%; Pred. No. 4.3e-239;
Matches 1243; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 944 GCGGTGGGCAGTGGCTGTGTGTGGAGGACCTAGGGACCCTTTCACACAGACTCTGTTGTTTG 1003
||| | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5 GCGTGTGGCAGTGGCTGTGTGTGGAGGACCTAGGGACCCTTTCACACAGACTCTGTTGTTTG 64

Qy 944 ATACTAGTTCCTGGAGGACCAGGGGTCTCCACCCAGCTACTCCCCCT-CTTTTCAGTTTT 1002
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||



```
Qy    2064 TTCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTTC 2109
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1205 TTCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTTC 1250
```

RESULT 7
AAS27110
ID AAS27110 standard; cDNA; 1247 BP.
XX
AC AAS27110;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 145.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial, antiinflammatory; cancer;
KW immune system disorder, rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection, infection; hepatitis C, blood disorder;
KW sickle cell anaemia, hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease, Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory, wound healing; endocrine, Addison's disease;
KW reproductive system; gastrointestinal, liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135
PR 07-JUL-2000; 2000US-0216647
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496
PR 14-JUL-2000; 2000US-0218290
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447
PR 14-AUG-2000; 2000US-0225757
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759
PR 18-AUG-2000; 2000US-0226279
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868
PR 23-AUG-2000; 2000US-0227009
PR 30-AUG-2000; 2000US-0228024
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.

```
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465460/50.
DR   P-PSDB; AAU17193.
XX
PT   Novel polypeptides useful for diagnosing, treating, preventing and/or
PT   prognosing disorders related to the proteins, including cancers, immune
PT   disorders and neuronal disorders
XX
PS   Claim 1; SEQ ID No 145; 880pp; English.
XX
CC   The invention relates to novel isolated polypeptides (I), and
CC   polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC   diagnosing, preventing and treating diseases including immune system
CC   disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC   disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC   transplant rejections and graft versus host disease, infectious diseases
CC   (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC   other blood-related disorders (sickle cell anaemia), myeloproliferative
CC   disorders, primary haematopoietic disorders, hyperproliferative
CC   disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC   abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC   disorders (e.g. glomerulonephritis), cardiovascular disorders
CC   (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC   wound healing, epithelial cell proliferation, endocrine disorders (e.g
CC   Addison's disease), reproductive system disorders, gastrointestinal
CC   disorder (inflammatory disorders), liver disorders (cirrhosis),
CC   as stimulators of B-cell responsiveness to pathogens, activators of
CC   T-cells, to induce higher affinity antibodies, and as a means to induce
CC   tumour proliferation in pathologies e.g. acquired immune deficiency
CC   syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC   pathway protein coding sequences and PCR primers of the invention.
XX

  Query Match             29.5%;  Score 1232.2;  DB 22;  Length 1247;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-237;
  Matches 1231;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    [illegible];

Qy     872 ACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGG[illegible] 931
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGG[illegible] 60

Qy     932 [illegible] 991
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      61 CGTGGAGGACTAGGGGTTCCACCCAGCTACTCGCGCCCTCTGGCAGCGTGTGCTTCTC 120

Qy    992 CTACACCGGCACGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGGGAGAA 1051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTACACCGGCACGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGGGAGAA 180

Qy   1052 CTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTC 1111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTC 240

Qy   1112 CGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGG 1171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGG 300

Qy   1172 CGTGGAAGCGGACTTGGATTCTCTACCACCACCGAAGACAGCGTCAAGAAGCGCATCGT 1231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTTGGAGGTGGACCTGGATTCTCTACCACCACCGAAGACAGCGTCAAGAAGCGCATCGT 360

Qy   1232 GGTGGCCGTTGGGGACAACTGGGCCAATTACTTCTCCCGCTTCTTTCCTGTCTCGGGCGA 1291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTGGCCGTTGGGGACAACTGGGCCAATTACTTCTCCCGCTTCTTTCCTGTCTCGGGCGA 420

Qy   1292 GAGTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTCAAGGT 1351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAGTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTCAAGGT 480

Qy   1352 GACCCAAGGCCCCCGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 1411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GACCCAAGGCCCCCGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 540

Qy   1412 GGAGGTGGTGGGTGTGGAGTGCCGGGGCGGCGTCCACCCTGGAGCTGTCACTGAAGAGCGA 1471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGAGGTGGTGGGTGTGGAATGCCGGGCGGCGTCCACCCTGGAGCTGTCACTGAAGAGCGA 600

Qy   1472 GCAGCTGGTGCTGCACACAGCCCGGCAAGGGCATCGAGGCGCTGGTTGAGCTATTCCT 1531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCAGCTGGTGCTGCACACAGCCCGGCAAGGGCATCGAGGCGCTGGTTGAGCTATTCCT 660

Qy   1532 GAATGAGCTTAAGAAGGATCTCGGCTATGTCATCGGCCTGCGCAGCTACATCACTGACAA 1591
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAACGAGCTTAAGAAGGATCTCGGCTATGTCATCGGCCTGCGCAGCTACATCACTGACAA 720

Qy   1592 CTGCAGCCTCCTCAGCTTCCACCTTGGGACCTCATCAAGCTGCTGCCGGTGGCCACCCT 1651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTGCAGCCTCCTCAGCTTCCACCTTGGGACCTCATCAAGCTGCTGCCGGTGGCCACCCT 780

Qy   1652 GGAGCCAGAATGGCAGTTTGGTCTGCGGGGGGCCGTTCCGGACTCTTTCCTGCCGACAT 1711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGAGCCAGGCTGGCAGTTTGGTCTGCGGGGGGCCGTTCCGGACTCTTTCCTGCCGACAT 840

Qy   1712 AGTGCAGCCGGCTGCCGCTGCCACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 1771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGTGCAGCCGGCTGCCGCTGCCACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 900

Qy   1772 CAAGGGTCAGCTGTCCAACGGGAACCAGGGCTGGCTGGGTGGGACAGGGCCCTCAGAGGT 1831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAAGGGTCAGCTGTCCAACGGGAACCAGGGCTGGCTGGGTGGGACAGGGCCCTCAGAGGT 960

Qy   1832 GAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGG 1891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGG 1020

Qy   1892 GGTTTGACCACTCCCGACGTTCTCATCCGTGTGCACCACCCTGCTGTCCGTCTCCTGTGG 1951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GGTTTGACCACTCCCGACGTTCTCATCCGTGTGCACCACCCTGCTGTCCGTCTCCTGTGG 1080

Qy   1952 CTGCCCTTCTTCCTCCTCCTGACCGCTGCCTTGTCTCTCCAGCAAGACTGTGCACTCCT 2011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTGCCCTTCTTCCTCCTCCTGACCGCTGCCTTGTCTCTCCAGCAAGACTGTGCACTCCT 1140

Qy   2012 TGCAGGCAGGGGCTGGGCTGGATGCTGCTCTTGTGTCCCACGTGGTACTTAGTTCAAGGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGCAGGCAGGGGCTGGGCTGGATGCTGCTCTTGTGTCCCACGTGGTACTTAGTTCAAGGC 1200

Qy   2072 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 2108
          |||||||||||||||||||||||||||||||||||||
Db   1201 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 1237


RESULT 8
ABK43853
ID   ABK43853 standard; cDNA; 1247 BP.
XX
AC   ABK43853;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding novel central nervous system protein #433.
XX
KW   central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder, neoplasm, cardiovascular disorder;
KW   cardiac arrest, cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS, ocular disorder;
KW   acquired immunodeficiency virus; dysphagia, gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder, testicular feminisation;
KW   endocrine disorder, diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder, kidney failure; blood disorder;
KW   myocardial infarction; wound healing, cell proliferation; skin aging;
KW   food additive; food preservative, gene therapy, gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155318-A2.
XX
PD   02 AUG 2001
XX
PF   17-JAN-2001; 2001WO-US01332.
XX
PP   31-JAN-2000; 2000US-0179065.
PP   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   14-AUG-2000; 2000US-0226279.
PR   14-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   23-AUG-2000; 2000US-0227180.
PR   28-AUG-2000; 2000US-0227994.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
```

PR 21 SEP 2000; 2000US-0229344.
PR 21 SEP 2000; 2000US-0229345.
PR 25 SEP 2000; 2000US-0229509.
PR 25 SEP 2000; 2000US-0229513.
PR 26 SEP 2000; 2000US-0229437.
PR 26 SEP 2000; 2000US-0229438.
PR 08-SEP-2000; 2000US-0231240.
PR 08-SEP-2000; 2000US-0231241.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 13-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 25-OCT-2000; 2000US-0241785.
PR 25-OCT-2000; 2000US-0241786.
PR 25-OCT-2000; 2000US-0241787.
PR 25-OCT-2000; 2000US-0241808.
PR 25-OCT-2000; 2000US-0241809.
PR 26-OCT-2000; 2000US-0241826.
PR 31-OCT-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR P-PSDB; AAU87523.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives
XX
PS Claim 1; SEQ ID No 443; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical condition and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dyspepsia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 29.5%; Score 1242.2; DB 13; Length 1247;
Best Local Similarity 99.5%; Pred. No. 2.7e-247;
Matches 1231; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 872 ACTGCCTCTGGCCGAGGACCCAGGGACCCTTCAGCAGA[illegible] 914
Db 1 ACTGCCTCTGGCCGAGGACCCAGGGACCCTTCAGCAGAGCTTTTTTTTTACACAGCC 60

Qy 915 [illegible] 964
Db 61 [illegible] 120

```
Db      61 CGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGGGGCCTCTGGCAGCGTGTGCTTCTC 120
Qy     992 CTACACCGGGCACGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGGGAGAA 1051
Db     121 CTACACCGGGCACGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGGGAGAA 180
Qy    1052 CTTCAGCCATCCCTACTACTTAGGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTC 1111
Db     181 CTTCAGCCATCCCTACTACTTAGGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTC 240
Qy    1112 CGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGGAAAATGAAAGACCTGCTGGGAGG 1171
Db     241 CGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGGAAAATGAAAGACCTGCTGGGAGG 300
Qy    1172 CTTGGAGGTGGACCTGGATTTCTCAGCACCACCGAAGACAGCGTCAAGAAGCGCATCGT 1231
Db     301 CTTGGAGGTGGACCTGGATTTCTCAGCACCACCGAAGACAGCGTCAAGAAGCGCATCGT 360
Qy    1232 GGTGGCGGGCTCGGGACAACTGGCGCAATTACTTCTCCCGTTCTTTCCTGTCTCGGGCGA 1291
Db     361 GGTGGCGGGCTCGGGACAACTGGCGCAATTACTTCTCCCGCTTCTTTCCTGTCTCGGGCGA 420
Qy    1292 GAGTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTCAAGGT 1351
Db     421 GAGTGGCAGCGACCTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTCAAGGT 480
Qy    1352 GACCCAAGGCCCCGGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 1411
Db     481 GACCCAAGGCCCCGGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 540
Qy    1412 GGAGGTGCTGGGTGTGGAATGCCGGGGCCGCTCCAGCCTGGAGCTGTCACTGAAGAGCGA 1471
Db     541 GGAGGTGCTGGGTGTGGAATGCCGGGGCCGCTCCAGCCTGGAGCTGTCACTGAAGAGCGA 600
Qy    1472 GCAGCTGGTGCTGCACACAGCCCGGCAAGGGCCATCGAGGCGCTGGTTGAGCTATTCCT 1531
Db     601 GCAGCTGGTGCTGCACACAGCCCGGCAAGGGCCATCGAGGCGCTGGTTGAGCTATTCCT 660
Qy    1532 GAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAA 1591
Db     661 GAACGAGCTTAAGAAGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAA 720
Qy    1592 CTGTAGCCTCCCTAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGCCGGTGGCCACCCT 1651
Db     721 CTGCAGCCTCCCTAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGCCGGTGGCCACCCT 780
Qy    1652 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACAT 1711
Db     781 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACAT 840
Qy    1712 AGTGCAGCGGGCCTGCCGCTCCGGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 1771
Db     841 AGTGCAGCGGGCCTGCCGCTCCGGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 900
Qy    1772 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGT 1831
Db     901 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGT 960
Qy    1832 GAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGG 1891
Db     961 GAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGG 1020
Qy    1892 GGTTTGACTACTCCCCAGGCTGCCATGCGGTGGGACCACCCTGCTGTCCGTCTCCTGTGG 1951
Db    1021 GGTTTGACTACTCCCCAGGCTGCCATGCGGTGGGACCACCCTGCTGTCCGTCTCCTGTGG 1080
Qy    1952 CTGCCGCTGTGCCCTGCTCATGGTCGCCTTCTCTCTCCAGCAAGACTGTGCACTCCT 2011
Db    1081 CTGCCGCTGTGCCCTGCTCATGGTCGCCTTGTCTCTCCAG-AAGACTGTGCACTCCT 1140
Qy    2012 TGCAGGCAGGGGCTGGGCTGGATGCTGCTTGTGTCCCACGTGGTACTTAGTTCAAGGC 2071
Db    1141 TGCAGGCAGGGGCTGGGCTGGATGCTGCTTGTGTCCCACGTGGTACTTAGTTCAAGGC 1200

Qy    2072 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 2108
Db    1201 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 1237
```

RESULT 9
AAS33274
ID AAS33274 standard; cDNA; 1346 BP
XX
AC AAS33274;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 233.
XX
KW Immunomodulatory, human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic, cardiant, vascular;
KW cerebroprotective, thrombolytic; antimicrobial, ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human, cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder, infection;
KW Gaucher's disease, neurological disease, cerebrovascular disorder;
KW thrombosis; wound healing; ss
XX
OS Homo sapiens.
XX
PN WO200155326 A2
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001, 2001WO-US01347
XX
PR 31-JAN-2000, 2000US-0179065.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-451931/48.
DR P-PSDB; AAU20565
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
XX
PS Claim 1; SEQ ID NO 233; 752pp. English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacterial, viruses and
CC fungi and ocular disorders (e.g. corneal infection). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture

```
CC   of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC   coding sequences, PCR primers, and related sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1346 BP; 294 A; 410 C; 407 G; 229 T; 6 other;

  Query Match             29.4%;  Score 1225.8;  DB 22;  Length 1346;
  Best Local Similarity   98.5%;  Pred. No. 5.3e-236;
  Matches 1235;  Conservative    4;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    2197 GCTGACAGGTTTCTTGTGGATGCCAGGGCCCTGCCCACCCTTGGAGCCAGGCACACAGT 2256
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1 GTGGCACCAGGTTTCTTGTGGCTCCCAGGGCCCTGCCCACCCTTGGAGCCAGGCACACAGT 60

Qy    2257 GACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCAC 2316
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCAC 120

Qy    2317 AGCTACACCATGCAGGAATTCGCCGGGGTTACTTCCGGAGGTCCCAGGCCTTGCTGGGC 2376
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGCTACACCATGCAGGAATTCGCCGGGGTTACTTCCGGAGGTCCCAGGCCTTGCTGGGC 180

Qy    2377 CTGACTGATGGAGGTTCCTCCAGGAAAGGACAGGACAGCCTGGTGCAGTACCAAGGCT 2436
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAGACTGATGGAGGTTCCTCCAGGAAAGGACAGGACAGCCTGGTGCAGTACCAAGGCT 240

Qy    2437 TTGATCAGGAGTCGCTCCTCAGCCTCAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGC 2496
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTGATCAGGAGTCGCTCCTCAGCCTCAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGC 300

Qy    2497 TTGTTGGCCCCTGATGCGGTTTATGGGCAGCAGTCCAAGCCCCGGGGAAGGATGAGAT 2556
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTGTTAGCCCCTGATGCGGTTTATGGGCAGCAGTCCAAGCCCCGGGGAAGGATGAGAT 360

Qy    2557 GGGTTTTTTTATAGAACTGCTCAAGCCTCTGCCAGCAGCAGAAGCTGAGGGATGAGATTTA 2616
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGATTGTTCTATGAACTGCTGAAGCTGTTGCCAGCAGCAGAAGCTGAGGGATGAGATTTA 419

Qy    2617 TTGTCAGATTATCAAGCAGGTCACAGGACACCCCGGGCCGGAACACTGCACTCGAGGCTG 2676
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 TTGTCAGATTATCAAGCAGGTCACAGGACACCCCGGGCCGGAACACTGCACTCGAGGCTG 479

Qy    2677 GAGCTTCTCAGCCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGATGCCCTACCT 2736
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 GAGTTTCTCAGCCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGATGCCCTACCT 539

Qy    2737 GACAAGTTTCTGCAGGATTCAGGCCCCAGCCAAGAGCTGGCCCCGAGCAGCAGGAGCA 2796
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GACAAGTTTCTGCAGGATTCAGGCCCCAGCCAAGAGCTGGCCCCGAGCAGCAGGAGCA 599

Qy    2797 CCTCCAGCCGCACAGTCAAATATGGGGGGCGCCGCGGATGCCCCCACCGGGTGAAATGAA 2856
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 CCTCCAGCCGCACAGTCAAATATGGGGGGCGCCGCGGATGCCCCCACCGGGTGAAATGAA 659

Qy    2857 GGTTTCTTAAAGGACAAGCGGATTGGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGA 2916
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 GGTTTCTTAAAGGACAAGCGGATTGGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGA 719

Qy    2917 TTATACCACAAGAATATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCG 2976
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 TTATACCACAAGAATATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCG 779

Qy    2977 GCAAATGGGTATCACCGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGA 3036
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 GCAAATGGGTATCACCGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGA 839

Qy    3037 AAAGAGTCAGCTGGTGCGCCCCTGCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGA 3096
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 GAAGAGTCAGCTGGTGCGCCCCTGCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGA 899

Qy    3097 GCAGATTTAGCCTGCAGTGCGGTGGGTTCACTGGAGACCCTATTGCATTCGATAA 3150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GCAGGACGTGAGCCTGCACAGCGTGGGTTCACTGGAGACCCTATTGCATTCGATAA 953

Qy    3151 CTCCACCTACATCATCACCATCATCAGCAAGCAGTGACTGTGAGACTATTTCAGGGAAGCT 3210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     954 CTCCACCTACATCATCACCATCATCAGCAAGCAGTGACTGTGAGACTATTTCAGGGAAGCT 1013

Qy    3211 GCCAGTCAGGGCCAAGGCAGACGGGTAGCTTGCCAGGTTGGCCTTAGCAGCTTCATT 3270
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 GCCAGTCAGGGCCAAGGCAGACGGGTAGCTTGCCAGGTTGGCCTTAGCAGCTTCATT 1073

Qy    3271 CAAGGCCAACAGCAATACCCCCTCAGGGCAGGACCTGCTACTTACCTGCCAAAGCAGCT 3330
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1074 CAAGGCCAACAGCAATACCCCCTCAGGGCAGGACCTGCTACTTACCTGCCAAAGCAGCT 1133

Qy    3331 GCAACGGCAGGTGAACACGGCCTCCATCAAGAACTCAATGGCTAACAACTGAGACCCT 3390
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1134 GCAACGGCAGGTGAACACGGCCTCCATCAAGAACTCAATGGCTAACAACTGAGACCCT 1193

Qy    3391 GGAAGGACACAGCCCCCAGGAAGCACAGATCAGCTTCATTG 3431
           |||||||||||||||||||||||||||||||||||||||||
Db    1194 GGAAGGACACAGCCCCCAGGAAGCACAGATCAGCTTCATTG 1234


RESULT 10
AAS13820
ID   AAS13820 standard; cDNA; 4375 BP.
XX
AC   AAS13820;
XX
DT   18-DEC-2001  (first entry)
XX
DE   cDNA #7 encoding a novel protein related to MMP.
XX
KW   Mouse; oestrogen regulated unconventional myosin related protein; MRP;
KW   auditory, antiatherosclerotic, osteopathic; neuroprotective; nootropic;
KW   antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW   osteoporosis; breast cancer; cardiovascular disease; deafness; CNS;
KW   Alzheimer's disease; Parkinson's disease; ataxia; stroke; immunopathy.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             125..4138
FT                   /*tag= a
FT                   /product= "mMRP"
FT                   /note= "Oestrogen regulated unconventional myosin
FT                         related protein"
XX
PN   WO200168866-A2.
XX
PD   20-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-US08060.
XX
PR   10-MAR-2000; 2000US-188488P.
PR   08-MAR-2001; 2001US-0803126.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Brooks AR,  Craig CG,  Kalmyl GM;
XX
DR   WPI; 2001-602747/68.
DR   P-PSDB; AAU07399.
XX
PT   Novel oestrogen-regulated unconventional myosin related protein and
PT   polynucleotide useful for altering effects of oestrogen in mammalian cells to
PT   treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS   Claim 4; Fig 3; 108pp; English.
XX
CC   The invention relates to an isolated oestrogen regulated unconventional
CC   myosin-related protein (MRP). Modulating MRP is useful for modulating the
```

C effects of oestrogen mediated by oestrogen receptor alpha or beta in a
C mammalian cell. When the level of expression or activity of MRP is
C increased, the development of atherosclerosis or osteoporosis is
C inhibited and when the level is decreased the development of breast
C cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
C useful for detecting the presence of oestrogen signaling in a mammalian
C cell, by detecting the expression or activity of MRP or the level of MRP
C mRNA in the cell. The presence of oestrogen signaling is used to
C determine the responsiveness of the cell to the oestrogen or the tissue-
C specific distribution of oestrogen signaling in a mammal. MRP is useful
C as a marker gene for the investigation of tissue-specific and oestrogen
C receptor-specific ligands such as oestrogen receptor antagonists and
C agonists. Modulators, recombinant forms, derivatives, variants or
C fragments of MRP are useful for enhancing or inhibiting oestrogen
C signaling in cells and for treatment of breast cancer or other cancers
C and for preventing or treating wide variety of oestrogen-related
C diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
C disease or Parkinson's disease. Modulation of MRP gene is also useful for
C preventing and/or treating hearing loss and deafness. MRP polypeptides
C are useful for generating monoclonal and polyclonal antibodies useful for
C identifying oestrogen responsive cells, particularly cells responsive to
C oestrogen through an oestrogen receptor alpha or beta, to identify
C alleles, interspecies homologues and polymorphic variants and in
C immunoassays to detect MRP. The present sequence represents the coding
C sequence of mouse MRP variant #2 as described in the invention.
XX
SQ Sequence 4375 BP; 986 A; 1368 C, 1215 G, 825 T; 0 other;

Query Match 22.7%; Score 946.4; DB 22; Length 4375;
Best Local Similarity 72.9%; Pred. No. 6.6e-180;
Matches 1322; Conservative 0; Mismatches 416; Indels 76; Gaps 5;

Qy 4 CAGCAGGA...GGAGGGTCGGTCCCAGGCCTCACCCTCAGCCGTCACCTCC 63
Db 2297 CAGCACCAAGGTCAGAGGTTGTGGTCAGTTGAGGCTCCACCCTCAAGCACTGCTCCC 2356

Qy 64 AAGCCCATGAAG...ACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAGGGT 123
Db 2357 AAGGTGAAGAAGCTGTGGCCCCCAAGAGAAGCCAGAGAGTAACCTAGAGCCTTCGGGT 2416

Qy 124 TGCTGCTTGAG...GGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
Db 2417 GTTGGTTGAGAGAGGACACCCAGAGGAAGCTGAAAGCAAGCCTCAGCGCCCCAAGAGC 2476

Qy 181 TTCCATCAGAAACGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
Db 2477 TTCCAACAGAAACGGGACTATTTCCAGAAGATGGGGTAAGATCCGATCAGAGTGAAGACG 2536

Qy 241 ATGAA...AAGGTCCATATCCCCCAGGGGAAGCGCAGGAGGAGGAGGAGGAG 300
Db 2537 GTGAAA...TCCAGCAAGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAG 2596

Qy 301 GAG...AAGTGCCAGCAGGAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360
Db 2597 GAGGATTAGA...GAGTTG----------TCCCCTCCTCCT 2629

Qy 361 ...ATGGTGGGGCATCTGTAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
Db 2630 GGT...TAAGAAGCCGCTGAAGGCAAGCAGGCCCAAAGCCGTAAAGGAA 2689

Qy 421 GGTTAGGTTGAGCCAGCCAAGAG----------AC 446
Db 2690 GATAATTAGAGAGGGGG...AGTAGTAATTACCGACGCAGGGGAGGATCCCCCGGTGCAC 2749

Qy 447 AGGGGCCAAGGGCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTGCGCAGTCA 506
Db 2750 AGTTTAAATTCGGTA...AAAGCAGCAGGGTAGCGCAGTGCAGAGCTCC 2809

Qy 507 GAGTCCAAGCCCAAGGGGCCAACCCAGCAGGGAAATTGGCAACATCATCCGCATGTAC 566
Db 2810 AACTCGGCAGCTCCACGGCGGCAACCCAGCAGGGAAATCCGAAACATCATCCGAATGTAC 2869

Qy 567 CAGAGCGGCTGGTGCCGTGCAGCCATCCAGGCCTCCCAAAGCTTTC 626

Db 2870 CAGAGCCGTCCAGGGCCTGTGGCTGTGCCCGTACAACCCACCAGGCCCATCAAAACTTTT 2929

Qy 627 CTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGCTGGTATCAAGGTGGCAC 686
Db 2930 CAGAAGAAAAATGACCCTAAGGATGAGGCTTTGGCTAAGTTAGGGATAAATGGCGTGCAC 2989

Qy 687 TCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGAAAGGCCCCCCCCAGCTGTGGCTCCT 746
Db 2990 TTGCCCCTATCGA-- CATCGCCTAACCAAGGGAAGAGCTCTCCACCGGCTGTAGTTCCT 3046

Qy 747 CGACCCAAGGCCCGGCTACAGCTTGGGCCTCTAGCTCCATCAAGGAAAAGCAGGGGCCC 806
Db 3047 CGACCTAAGGCTCGACCTGTCTTGAGCCTTCCCTATCCATCCAGGAAAAGCAGGCACCC 3106

Qy 807 CTTCTGGACCTGTTTGGC--------- CAGAAGCTGCCTATTGCCCATACACCCCACCT 857
Db 3107 CTTCGGGACTTGTTTGGCCCATGTAGTCCAAACCCACCTACAGCTCCAGCACCCCCGCCT 3166

Qy 858 CCACCAGCACTCCCACCGCCTCTGTCTGGGAGCCCAAGACCCCTTCAGTGGAGTCTCAT 917
Db 3167 CCACCAGCACTCCCACCGCCTCTGTCTGGGAGCCCAAGACCCCTTCAGTGGAGTCTCAT 3226

Qy 918 TGCTTGACACAGCCCGTGGAGGACCAGGGGTCTCTGA...AGGTAGTCCGGGCCTCTGGC 977
Db 3227 GCCTTGACAGAGCCCATGGAGGACAAGAACATCTCCA...AAAGCCTTTTT...TTTC 3286

Qy 978 AGCGTGTGCTTCTCCTACACCGGCAGGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTC 1037
Db 3287 AGTGTGTGCTTCTCCTATGCCAATGCACCCTGGAAGTTGTTCTTA...AAGGAGGTGTTC 3346

Qy 1038 TACCCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTA 1097
Db 3347 TACCCCCGGGAGAACTTCAGTCATCCATACTGCCTCAGTCTCCTCTGC...AGCAGATCCTG 3406

Qy 1098 CGGGACACCTTCTCCGAGTCCTGTATCCGGATTTCCCAGAAT...A...AAAATGAAA 1157
Db 3407 CGGGACACCTTCACAGAGTCCTGCACCCGGATCTCACAGGAT...A...AAAAGGAA 3466

Qy 1158 GACCTGCTGGGAGGCTTGGAGGTGGA...GGATTCTCCCA...A...AGCAGGTC 1217
Db 3467 GGCCTTCTGGGAGACTTGGAGGTGAGTCTGGAGACCTTTA...A...AGCAGCATC 3526

Qy 1218 AAGAAGCGCATCGTGGTGGCCGGTGGGACAAGT...AATTA...TCCCGCTTCTTT 1277
Db 3527 AAAAAAGGCATCGTGGTCGCTGGTGGGACAATT...AATTA...TCCCACGTGGGCTG 3586

Qy 1278 CCTGTCTCGGGCGAGAGTGGCAG...GTCTTA...CCCGGACTG 3646
Db 3587 CCAGTCTCGGGTGAGAGTGGCAGCGATGTACAGCTGCT...GTCCAGGGGACTG 3646

Qy 1338 CGACTGCTCAAGGTGACCCAAGGCCCCGGGCTCCGGCC...AAGATTCCTGC 1397
Db 3647 CGGCTGCTGAAGGTGACCCAAAGCCCGAGCTTCCAGCT...AAGACTCTGT 3706

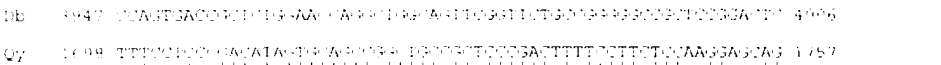
Qy 1398 TCATACAGCTTTGCGGAGGTGCTGGGTGTGGAGTGC...TCACCCTGGAGCTG 1457
Db 3707 TCCTACAGCTATGCTGAAGTCCTGACCGTGCAGT...AGATCCACCCTGGAGCTG 3766

Qy 1458 TCACTGAAGAGCGAGCAGCTGGTGCTG...AAGGCATG 3826
Db 3767 TCCTTGAAGAATGAGCAGCTGATACTGCACAC...AAGGCATG 3826

Qy 1518 GTTGAGCTATTCCTGAATGAGCTTAAGAAGGA...GGCAGC 3886
Db 3827 GTGGATCTATTTCTGAGTGAACTCAGGAAGGA...GGCAGC 3886

Qy 1578 TACATCACTGACAACTGCAGCCTCCTCAGCTTCCA...ACATC 3946
Db 3887 TACATCACCGATGACAATAGCCTCCTCAGTTTCCA...ACATC 3946

Qy 1638 CCGGTGGCCACCCTGGAGCCAGGCTGGCAGTTT...CCCGACTG 1697



```
RESULT 11
AAS27687
ID   AAS27687 standard; DNA; 24757 BP
XX
AC   AAS27687;
XX
DT   ?? NOV 2001  (first entry)
XX
DE   DNA encoding novel signal transduction pathway protein, Seq ID 174
XX
KW   Neuropsychiatric; cytostatic; human T cell; immunosuppressive; tumour;
KW   antiviral; anti-HIV; anti-retroviral; antiinflammatory; cancer;
KW   immune system disorder; rheumatoid arthritis; inflammatory condition;
KW   organ transplant rejection; infection; hepatitis C; blood disorder;
KW   sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW   cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW   reproductive system; gastrointestinal; liver disorder; AIDS; ds;
KW   immunodeficiency syndrome.
XX
OS   Homo sapiens
XX
PN   WO200147743-A1.
XX
PD   05-JUL-2001.
XX
PF   27-JAN-2001; 2001WO-US02431.
XX
PR   31-JAN-2000; 2000US-0179065
PR   04-FEB-2000; 2000US-0180628
PR   24-FEB-2000; 2000US-0184664
PR   07-MAR-2000; 2000US-0187552
PR   16-MAR-2000; 2000US-0189874
PR   17-MAR-2000; 2000US-0190376
PR   18-APR-2000; 2000US-0198127
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209762
PR   09-JUN-2000; 2000US-0210467.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216847
PR   07-JUL-2000; 2000US-0216886
PR   11-JUL-2000; 2000US-0217497
PR   11-JUL-2000; 2000US-0217496
PR   14-JUL-2000; 2000US-0218290
PR   [illegible]-JUL-2000; 2000US-0220863
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518
PR   14-AUG-2000; 2000US-0224519
PR   14-AUG-2000; 2000US-0224213
PR   14-AUG-2000; 2000US-0224214
PR   14-AUG-2000; 2000US-0224266
PR   14-AUG-2000; 2000US-0224267
PR   14-AUG-2000; 2000US-0224268.
PR   14-AUG-2000; 2000US-0224776
PR   14-AUG-2000; 2000US-0224447
PR   14-AUG-2000; 2000US-0224767
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759
PR   18-AUG-2000; 2000US-0226218.
PR   22-AUG-2000; 2000US-0226881
PR   22-AUG-2000; 2000US-0226868
PR   31-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0231080.
PR   08-SEP-2000; 2000US-0231081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236902.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240860.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0248207.
PR   17-NOV-2000; 2000US-0248208.
```

R 17-NOV-2000; 2000US-0249209.
R 17-NOV-2000; 2000US-0249210.
R 17-NOV-2000; 2000US-0249211.
R 17-NOV-2000; 2000US-0249212.
R 17-NOV-2000; 2000US-0249213.
R 17-NOV-2000; 2000US-0249214.
R 17-NOV-2000; 2000US-0249215.
R 17-NOV-2000; 2000US-0249216.
R 17-NOV-2000; 2000US-0249217.
R 17-NOV-2000; 2000US-0249218.
R 17-NOV-2000; 2000US-0249244.
R 17-NOV-2000; 2000US-0249245.
R 17-NOV-2000; 2000US-0249264.
R 17-NOV-2000; 2000US-0249265.
R 17-NOV-2000; 2000US-0249267.
R 17-NOV-2000; 2000US-0249299.
R 17-NOV-2000; 2000US-0249300.
R 01-DEC-2000; 2000US-0250160.
R 01-DEC-2000; 2000US-0250391.
R 05-DEC-2000; 2000US-0251040.
R 05-DEC-2000; 2000US-0251089.
R 05-DEC-2000; 2000US-0256719.
R 06-DEC-2000; 2000US-0251479.
R 08-DEC-2000; 2000US-0251866.
R 08-DEC-2000; 2000US-0251868.
R 08-DEC-2000; 2000US-0251869.
R 08-DEC-2000; 2000US-0251989.
R 08-DEC-2000; 2000US-0251990.
R 11-DEC-2000; 2000US-0254787.
R 05-JAN-2001; 2001US-0253678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT diagnosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 1347; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX
SQ Sequence 24757 BP; 5183 A; 7437 C; 7328 G; 4809 T; 0 other;

Query Match 17.1%; Score 715; DB 22; Length 24757;
Best Local Similarity 99.3%; Pred. No. 2.3e-133;
Matches 719; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1656 CCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTG 1715
Db 10170 CCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTG 10111

Qy 1716 CAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAG 1775
Db 10110 CAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAG 10051

Qy 1776 GGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGG 1835
Db 10050 GGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGG 9991

Qy 1836 AAGATGGGAGAGGGACAAGCAGAGGCAAGGGCT[illegible]GAGACTGAGGAAGGAAAGGGGTT 1895
Db 9990 AAGATGGGAGAGGGACAAGCAGAGGCAAGGGCT[illegible]GAGACTGAGGAAGGAAAGGGGTT 9931

Qy 1896 TGACCACTCCCGAGGCTGCCATGGGGTGGGACCA[illegible] 1955
Db 9930 TGACCACTCCCGAGGCTGCCATG[illegible] 9871

Qy 1956 CCCTCTGCCCGCTCCTGATGGCTCGCCTTGTCTCCTCCAGCAAGACTGTGCACTCCTTGCA 2015
Db 9870 CCCTCTGCCCGCTCCTGATGGCTCGCCTTGTCTCCTCCAGCAAGACTGTGCACTCCTTGCA 9811

Qy 2016 GGCAGGGGCTGGGCTGGATGCTGCTCTTGTGTCCCACCTGGTACTTAGTTCAAGGCTGCC 2075
Db 9810 GGCAGGGGCTGGGCTGGATGCTGCTCTTGTGTCCCACCTGGTACTTAGTTCAAGGCTGCC 9751

Qy 2076 CCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTGGCTTCTGGTCTTGCTCCTTTGGT 2135
Db 9750 CCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTGGCTTCTGGTCTTGCTCCTTTGGT 9691

Qy 2136 GTCTGGCTGGGGAGGGATGGGGCTGGGGCAGGACCCCTGGGACAGGGCACTGGACACTCA 2195
Db 9690 GTCTGGCTGGGGAGGGATGGGGCTGGGGCAGGACCCCTGGGACAGGGCACTGGACACTCA 9631

Qy 2196 GGTGGCACCAGGTTTCTTGTGATCCCAGCGCCCTGCCCACCCTTGGAGCCAGGCACACAG 2255
Db 9630 GGTGGCACCAGGTTTCTTGTGATCCCAGCGCCCTGCCCACCCTTGGAGCCAGGCACACAG 9571

Qy 2256 TGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCA 2315
Db 9570 TGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCA 9511

Qy 2316 CAGCTACACCATGCAGGAATTCGCCCGGCGTTACTTCCGGAGGTCCCAGGCTTGTGAGT 9451
Db 9510 CAGCTACACCATGCAGGAATTCGCCCGGCGTTACTTCCGGAGGTCCCAGGCCTTGTGAGT 9451

Qy 2376 CCA 2378
Db 9450 GCA 9448

RESULT 12
AAS33481
ID AAS33481 standard, DNA, 24757 BP
XX
AC AAS33481;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 764.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; anemia;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease, Parkinson's disease, human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease, neurological disease, cerebrovascular disorder;
KW thrombosis; wound healing; ds.
XX
OS Homo sapiens.
XX
PN WO200155226-A2

```
XX
PD   02 AUG 2001.
XX
PF   17 JAN 2001; 2001WO-US01347.
XX
PR   31 JAN 2000; 2000US-0179065.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Disclosure; SEQ ID NO 764; 773pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example, immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot Marie-Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC   anemia and thrombosis), infections caused by bacteria, viruses and
CC   fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC   agonists, antagonists and antibodies can also be used to promote wound
CC   healing, maintain organs before transplantation, and support cell culture
CC   of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC   coding sequences, PCR primers, and related sequences of the invention
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 24787 BP; 4809 A; 7328 C; 7437 G; 5183 T; 0 other;

  Query Match             17.1%;  Score 715;  DB 20;  Length 24787;
  Best Local Similarity   99.3%;  Pred. No. 2.3e-133;
  Matches  719;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     1456 CCAGGCTGGCAGTTGGCTCTGCCCCCCGCCTTCCCGACTCTTTCTTGCCGACATAGTG 1515
Db    14588 CCAGGCTGGCAGTTGGCTCTGCCCCCCGCCTTCCCGACTCTTTCTTGCCGACATAGTG 14647

Qy     1516 GATCCGACTGCCCCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTCGCTGGCACAAG 1575
Db    14648 GATCCGACTGCCCCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTCGCTGGCACAAG 14707

Qy     1576 CGTCAGCTGTCCAAGGGGAACCAGGGCTGGCTGGGTGGGACATGGTCTAGAGGTGAGG 1635
Db    14708 CGTCAGCTGTCCAAGGGGAACCAGGGCTGGCTGGGTGGGACAGGGCTCAGAGGTGAGG 14767

Qy     1636 AACATGGAGAGGGACAAGTAGAGGCAAGGCTGGCTGAGACTGATAAGGAAAGGGTT 1690
Db    14768 AACATGGAGAGGGACAAGTAGAGGCAAGGCTGGCTGAGACTGATAAGGAAAGGGTT 14822

Qy     1896 TGACCACTCCCCGAGGCTGCCATGCGGTGGGACACACTGTGTCTCTGTGGGTCT 1949
Db    14823 TGACCACTCCCCGAGGCTGCCATGCGGTGGGACACACTGTGTCTCTGTGGGTCT 14876

Qy     1950 CCCTCTGCCCCGTCCTCATGGCTCCCTCTCCAAAAGACTCTGCACTCCTTGCA 2003
Db    14877 CCCTCTGCCCCGTCCTCATGGCTCCCTCTCCAAAAGACTCTGCACTCCTTGCA 14930

Qy     2004 GGCAGGGGCTGGGCTGGATGCTCTCTGTGTCCCACTTACCTAAGGTGAAGCTGCC 2063
Db    14931 GGCAGGGGCTGGGCTGGATGCTCTTGTGTCCCACTTACCTAAGGTGAAGCTGCC 14990

Qy     2064 CCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTGGCTTCCCCCCCCCCCCTTTGGT 2123
Db    14991 CCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTGGCTTCCCCCCCCCCCCTTTGGT 15050

Qy     2124 GTCTGGCTGGGGAGGGATGGGGCTGGGGAAGAACCTGGGACAGGGCACTGGACACTCA 2183
Db    15051 GTCTGGCTGGGGAGGGATGGGGCTGGGGAAGAACCTGGGACAGGGCACTGGACACTCA 15110

Qy     2184 GGCGGCACCAGGTTTCTTGTGACCCAGGCCCTGCCAGCCCCAGGCACAGCAG 2237
Db    15111 GGCGGCACCAGGTTTCTTGTGACCCAGGCCCTGCCAGCCCCAGGCACAGCAG 15164

Qy     2238 TGACGACTCGGAGGCCACACAGCCTCTCCTGTGGGCTATCTTTCTGTGGACTCCA 2297
Db    15165 TGACGACTCGGAGGCCACACAGCCTCTCCTGTGGGCTATCTTTCTGTGGACTCCA 15224

Qy     2298 CAGCTACACACATCCACGGAATTCCCCCCCCCTTACTTCCAGAGTCTCTGTGGG 2357
Db    15225 CAGCTACACACATCCACGGAATTCCCCCCCCCTTACTTCCAGAGTCTCTGTGGG 15284

Qy     2358 CCA 2360
Db    15285 CCA 15287

RESULT 13
AAS33480
ID   AAS33480 standard; DNA; 9703 BP.
XX
AC   AAS33480;
XX
DT   04-DEC-2001  (first entry)
XX
DE   DNA encoding human secreted protein, Seq ID No 764.
XX
KW   Immunomodulatory; human immunodeficiency virus; HIV; anaemia; arthritis;
KW   rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW   cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW   multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW   Gaucher's disease; neurological disease; cerebrovascular disorder;
KW   thrombosis; wound healing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155326-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01347.
XX
PR   31-JAN-2000; 2000US-0179065.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
```

```
XX
DS   Disclosure; SEQ ID No 763; 753pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC   angina and thrombosis), infections caused by bacteria, viruses and
CC   fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC   agonists, antagonists and antibodies can also be used to promote wound
CC   healing, maintain organs before transplantation, and support cell culture
CC   of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC   coding sequences, PCR primers, and related sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9703 BP; 1950 A; 2842 C; 3079 G; 1832 T; 0 other;

  Query Match           10.6%;  Score 442;  DB 22;  Length 9703;
  Best Local Similarity  90.1%;  Pred. No. 8.5e-79;
  Matches  466;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy    3650 GGCCTGGAAGTCAACTATGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGCTGCCA 3709
           |||| ||         | | ||    ||||   | |||          | | ||  |
Db    5413 GGCCGGGGCAGAGGCCAGGGCTGGCTGGGGAGCCCTCTTGTAACAGGCCTTTCTCTGGG 5472

Qy    3710 CAGGCCCAGGAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACT 3769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5473 CAGGCCCAGGAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACT 5532

Qy    3770 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCCTCC 3829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5533 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCCTCC 5592

Qy    3830 CCCGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 3889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5593 CCCGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 5652

Qy    3890 CCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAC 3949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5653 CCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAC 5712

Qy    3950 GTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGG 4009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5713 GTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGG 5772

Qy    4010 GCTGCAGGAACTCGGCTGGTGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGA 4069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5773 GCTGCAGGAACTCGGCTGGTGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGA 5832

Qy    4070 CCCCAGGCTCCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCT 4129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5833 CCCCAGGCTCCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCT 5892

Qy    4130 GAGGAAATAAAACTCCCTAAAAAAAA 4155
           ||||||||||||||||||  | ||||
Db    5893 GAGGAAATAAAACTCCCTGGAGAAAA 5918


RESULT 14
AAD07791
ID   AAD07791 standard; cDNA; 1044 BP.
XX
AC   AAD07791;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 8 cDNA clone HECNC95, SEQ ID NO:32.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             616..657
FT                   /*tag= a
FT                   /product= "Human secreted protein precursor"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200132687-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US29364.
XX
PR   29-OCT-1999;   99US-0162238.
PR   30 JUN-2000; 2000US-0215134
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Duan RD,  Fiscella M,  Ebner R;
XX
DR   WPI; 2001-398740/32
DR   P-PSDB; AAE03399.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition
XX
PS   Claim 1; Page 362; 398pp, English
XX
CC   AAD07770-AAD07794 represent cDNAs corresponding to 10 human secreted
CC   protein genes, and AAE03378-AAE03402 represent the proteins they encode.
CC   AAE03403-AAE03415 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   10 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
```

```
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, neurogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as food additives or preservatives, as pH buffering agents, etc.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein encoding cDNA of the invention.
XX
SQ   Sequence 1044 BP; 212 A; 290 C; 326 G; 146 T; 5 other;

  Query Match             7.7%;  Score 322;  DB 22,  Length 1044;
  Best Local Similarity   98.1%;  Pred. No. 2.3e-59;
  Matches  312;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 GTGGTGCAGTGAGCTGTGATCGCGCTTCTGCACTCCAGCCTGGGCAGCAGAGCGAGACTCCGTCTCAAAAA 70
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     220 GTGATGCAGTGAGCTGTGATCGCGCTTCTGCACTCCAGCCTGCACAGCTCAGAGCGTCAGACCCTCAAAAA 779

Qy      71 TCGAAAGCCCAGAAAGCCCCCCCACACCCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TCGAAAGCCCAGGAAAGCCCCCCCACACCCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 839

Qy     131 GATGGCTGCTGCTGAGGAGACCTCCGAGCAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 GATGGCTGCTGCTGAGGAGACCTCCGAGCAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 899

Qy     191 TTCCAGCTACAAAACGGAACTATTCCCAGAGGATGGGGGCAGCCACAGATCACAGTGAGGAG 250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 TTCCAGCTACAAAACGGAACTATTCCCAGAGGATGGGGGCAGCCACAGATCACAGTGAGGAG 959

Qy     251 ATGAAGCTCCCCGGGCCAAGGTCCACATCCCCCAGGGGGAAGCCCAGGAGGAGGAGGAGGAGAG 310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 ATGAAGCTCCCCGGGCCAAGGTCCACATCCCCCAGGGGGAAGCCCAGGAGGAGGAGGAGGAGAG 1019

Qy     311 GAGAGAGAGGAGGAGGAGGAGCAGGAGG 338
           |||||||||||||||||||||||||||||
Db    1020 GAGAGAGAGGAGGAGGAGGAGCAGGAGG 1047


RESULT 15
ABN95022/c
ID   ABN95022 standard; DNA; 311 BP.
XX
AC   ABN95022;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Gene #1122 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour, cytostatic, expression profile, disease state,
KW   disease progression, drug toxicity, drug efficacy, drug metabolism
XX
OS   Homo sapiens.
XX
PN   WO200228433-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE ) GENE LOGIC INC.
XX
PI   Horne D,  Alvarez C,  Perez-Da-Silva S,  Yuckley JG;
XX
DR   WPI; 2002-405113/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample -
XX
PS   Claim 1; SEQ ID NO 1122; 238pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN97455 ABN97456 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease status, disease progression,
CC   drug toxicity, drug efficacy, and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 311 BP; 44 A; 85 C; 104 G; 79 T; 0 other;

  Query Match             7.0%;  Score 292;  DB 24;  Length 311;
  Best Local Similarity   98.1%;  Pred. No. 3.9e-49;
  Matches  306;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy    3843 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTTAAAACTTCCCTCACAAC 3902
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTTAAAACTTCCCTCACAAC 252

Qy    3903 AGCCCAGCCGGGCCCACATGCAGCCATGAGGAAAAACTGCTATCACTCTCACTCAATCCTTC 3962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 AGCCCAGCCGGGCCCACATGCAGCCATGAGGAAAAACTGCTATCACTCTCACTCAATCCTTC 192

Qy    3963 AAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGCTGCAGAACTTCAGAA 4022
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 AAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGCTGCAGAACTTCAGAA 132

Qy    4023 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCAAGCCAGGTCCTGTGCA 4082
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCAAGCCAGGTCCTGTGCA 72

Qy    4083 CCAGGCCCCACATTAGCACAAGCCCAGCCATGGAGAAACAGCTGCTGAGGAAATAAAAC 4142
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 CCAGGCCCCACATTAGCACAAGCCCAGCCATGGAGAAACAGCTGCTGAGGAAATAAAAC 13

Qy    4143 TCCCTAAAAAAA 4154
           ||||| | |||
Db      12 TCCCTGGAGAAA 1


Search completed: January 9, 2003, 07:19:36
Job time : 987.225 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:23:47 ; Search time 4982.49 Seconds
(without alignments)
13567.518 Million cell updates/sec

Title: US-09-803-126-5
Perfect score: 4174
Sequence: 1 cggcgg...gaaaaaaaaaaaaaaaaaaa 4174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154966 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters. 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	874.8	21.0	969	14	BM804816	BM804816 AGENCOURT
2	783.2	18.8	920	12	BG680906	BG680906 602629729
3	754	18.1	1004	14	BM804831	BM804831 AGENCOURT
4	736.6	17.6	799	13	BI772324	BI772324 603056028
5	735.8	17.6	1070	13	BI766913	BI766913 603053164
6	711.2	17.0	1015	14	BM923763	BM923763 AGENCOURT
7	703.6	16.9	836	13	BG912704	BG912704 602807089
8	699.6	16.8	908	13	BI766411	BI766411 603052657
c 9	696	16.7	715	14	BQ574071	BQ574071 UI-H-BZ0-
c 10	683.2	16.4	689	14	BM977966	BM977966 UI-CF-EC1
11	665.8	16.0	728	14	BM720779	BM720779 UI-E-EC1
12	665	15.9	754	13	BM545531	BM545531 AGENCOURT
c 13	659.4	15.8	679	14	BM971659	BM971659 UI-CF-EC1
c 14	645.8	15.5	668	14	BM978287	BM978287 UI-CF-EC1
c 15	625.8	15.0	649	9	AI671062	AI671062 tc17f07.x
16	607.8	14.6	626	10	AW753302	AW753302 RC3-CT025
17	606.4	14.5	661	13	BI769659	BI769659 603055655
18	603	14.4	603	14	BM791149	BM791149 K-BST0071
19	592	14.2	623	13	BG949021	BG949021 QV1-HB003
c 20	585.4	14.0	676	10	AV683102	AV683102 AV683102
c 21	571.6	13.7	579	9	AI023330	AI023330 ...c01f08.x
22	571	13.7	877	12	BG426026	BG426026 602492704
c 23	564	13.5	567	9	AI825877	AI825877 td19g01.x
c 24	555.2	13.3	637	13	BI046691	BI046691 MR3-TN020
c 25	544	13.0	564	10	AW243448	AW243448 aw95g01.x
c 26	543.6	13.0	591	12	BF817068	BF817068 MR2-CI012
c 27	541.6	13.0	549	12	BF436798	BF436798 ...p04c12.x
c 28	533.2	12.8	538	14	BM996595	BM996595 UI-H-DT0
c 29	519.4	12.4	594	10	AW854251	AW854251 RC3-CT025
30	508.6	12.2	523	14	BQ956737	BQ956737 AGENCOURT
c 31	501.2	12.0	510	9	AI990947	AI990947 ...c08f08.x
32	487.4	11.7	881	13	BG968181	BG968181 602835662
33	484.4	11.6	722	13	BI772088	BI772088 603059167
34	482.2	11.6	530	13	BI007177	BI007177 QV1-BT007
35	479.4	11.5	669	12	BF334334	BF334334 RC3-CT025
c 36	478.6	11.5	503	14	BM672910	BM672910 UI-H-CQ0
c 37	475.2	11.4	511	10	AW081367	AW081367 xc41d09.x
38	467	11.2	478	12	BF805323	BF805323 QV1-CI017
39	466.2	11.2	567	12	BG877603	BG877603 PM2-HT052
c 40	463	11.1	471	10	AW361379	AW361379 RC3-CT025
c 41	462.4	11.1	481	12	BF116088	BF116088 ...n23f1.x
c 42	460.4	11.0	473	13	BM150946	BM150946 TCBAP1D12
43	460.2	11.0	485	14	BQ374676	BQ374676 RC0-TN007
c 44	457.2	11.0	463	12	BF434203	BF434203 ...p01b01.x
45	452.2	10.8	457	10	AW391909	AW391909 QV4-CT023

ALIGNMENTS

RESULT 1
BM804816
LOCUS BM804816 969 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT_6494843 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725517 5', mRNA sequence.
ACCESSION BM804816
VERSION BM804816.1 GI:19121639
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 969)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12716 row: i column: 06
High quality sequence stop: 626
FEATURES Location/Qualifiers
source 1..969

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5726517"
                     /clone_lib="NIH_MGC_125"
                     /lab_host="DH10B"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
BASE COUNT      208 a    333 c    267 g    158 t      3 others
ORIGIN

  Query Match             21.0%;  Score 874.8;  DB 14;  Length 969;
  Best Local Similarity   97.6%;  Pred. No. 1.9e-114;
  Matches  947;  Conservative    0;  Mismatches   18;  Indels    5;  Gaps    5;

Qy        322 GAGGAGTAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCTCCTCCTCCCCCCATCGTGAAG 381
Db          1 GAGGAGTAAGAAGTGGAAACAAGAGCAGTGCCGTCCCCTCCTCCTCCCCCCATCGTGAAG 60

Qy        382 AAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAG 441
Db         61 AAGCCATTGAAGCAAGG-----CAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAG 120

Qy        442 GAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTGCGC 501
Db        121 GAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTGCGC 180

Qy        502 AGCTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGC 560
Db        181 AGCTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGC 240

Qy        561 ATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAA 620
Db        241 ATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAA 300

Qy        621 GCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGCTGGGTATCAACGGT 680
Db        301 GCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGCTGGGTATCAACGGT 360

Qy        681 GCCCACTCGTCCCCGGCCGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCCGGCCAGCTGTG 740
Db        361 GCCCACTCGTCCCCGGCCGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCCGGCCAGCTGTG 420

Qy        741 GCCCCTGACCCAAGCCCCCCTACAGCTTGGCCCCTCTAGCTCCATCAAGGAAAAGCAG 800
Db        421 GCCCCTGACCCAAGCCCCCCTACAGCTTGGCCCCTCTAGCTCCATCAAGGAAAAGCAG 480

Qy        801 GGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACACACCCCCACCTCCA 860
Db        481 GGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACACACCCCCACCTCCA 540

Qy        861 CCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGC 920
Db        541 CCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGC 600

Qy        921 TTGACACAGCCCGTGGAGGACCAGGGGGTCCCCACCCAGCTACTCGCCCCTCTGGCAGC 980
Db        601 TTGACACAGCCCGTGGAGGACCAGGGGGTCCCCACCCAGCTACTCGCCCCTCTGGCAGC 660

Qy        981 GTGTGCTTCTCCTACACCGGCACGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTAC 1040
Db        661 GTGTGCTTCTCCTACACCGGCACGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTAC 720

Qy       1041 CCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGG 1100
Db        721 CCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGG 780

Qy       1101 GACACCTTCTCCGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGAC 1160
Db        781 GACACCTTCTTCGAGT-CTGTATCCGGATTTCCCAGAATGA-----GAAAATGAAAGAA 839

Qy       1161 CTGCTGG-GAGGCTTGGAGGTGGACCTGGATTCTCTCACCACCACGAAGACAGCGTCAA 1219
Db        840 CTGCTGGNNAGGCTTGGAAGTGGACCTGGAATCTCTCACCACCACGAAGA-AGCGTCAG 898

Qy       1220 GAAGCCATCGTGGTGGCCCCTCGGGAAGAACCGGGCAATACTTCTCCCGCTTCTTTTC 1278
Db        899 AAAGCCCATCGTGGTGGCCCTCCGGACCACTC----CATCACTTCTCCCGCTTCCTTTC 958

Qy       1279 CTG 1281
Db        959 CTG 961


RESULT 2
BG680906
LOCUS       BG680906                 920 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602628729F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4753615 5',
            mRNA sequence.
ACCESSION   BG680906
VERSION     BG680906.1  GI:13912316
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10613  row: i  column: 08
            High quality sequence stop: 889.
FEATURES             Location/Qualifiers
     source          1..920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4753615"
                     /clone_lib="NCI_CGAP_Skn4"
                     /tissue_type="squamous cell carcinoma"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      211 a    284 c    264 g    161 t
ORIGIN

  Query Match             18.8%;  Score 783.2;  DB 12;  Length 920;
  Best Local Similarity   95.7%;  Pred. No. 1.7e-101;
  Matches  858;  Conservative    0;  Mismatches   33;  Indels    6;  Gaps    5;

Qy       2691 TTCTCACAGGCTTCTTCCCCCGTCGACCAGGCTGATGCCCTACCTGACCAAGCTTCTGC 2750
Db         13 TTTTAGTTGTCAACTTCCCCCGTCGACCAGGCTGATGCCCTACCTGACCAAGCTTCTGT 72

Qy       2751 AGGATTCAGGCCCCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCACCTCCAGCGCACAG 2810
Db         73 AGGATTCAGGCCCCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCACCTCCAGCGCACAG 132

Qy       2811 TCAAATATGGGGGGCCCGGCCGATGCCCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAG 2870
```

```
Db    133 TCAAATATGGGGGGCGCCGGGCGGATGCCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAG 192

Qy   2871 GACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATA 2930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 GACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATA 252

Qy   2931 TCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTATCA 2990
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 TCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTATCA 312

Qy   2991 CGGAGGCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGG 3050
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 CGGAGGCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGG 372

Qy   3051 TGCGGCCGCTGCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCC 3110
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 TGCGGCCGCTGCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCC 432

Qy   3111 TGCACAGCCGGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCA 3170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 TGCACAGCCGGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCA 492

Qy   3171 GCACCCACTACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAGCTGCCAGTCAGCGCCA 3230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493 GCACCCACTACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAGCTGCCAGTCAGCGCCA 552

Db    553 AGGAGACGGCTACTTCCAGGCTGGCCGCCCTGCAGCACCTCAGCAAGGCCAACAGGA 610

Qy   3231 ATACCCCCTCAGGGCAGGACCTGCTAGCTTACGTGCCAAAGCAGCTGCAACGGCAGGTGA 3290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATACCCCCTCAGGGCAGGACCTGCTAGCTTACGTGCCAAAGCAGCTGCAACGGCAGGTGA 670

Qy   3291 ACACGGCCTCCATCAAGAACCTGATGGGTCAGGAGCTGAGAACGGTGGAAGGACACCAGC 3350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    673 ACACGGCCTCCATCAAGAACCTGATGGGTCAGGAGCTGAGAACGGTGGAAGGACACCAGC 732

Qy   3351 ACAGGCCTCCATCAAGAACCTGATGGGTCAGGAGCT-GCCCCTCTTCGGCTAC 3403
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    673 ACAGGCCTCCATCAAGAACCTGATGGGTCAGGAGCTGAGAACGGTGGAAGGACACCAGC 732

Qy   3404 CCCAGGAAGCACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGGCCCCTCTTCGGCTAC 3463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    733 CCCAGGAAGCACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGGCCCCTCTTCGGCTAC 792

Qy   3464 ATGGTCTATGGGGTGCTGG-AGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGG 3522
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    793 ATGGTCTATGGGGTGCTGGAAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGG 852

Qy   3523 GGCTCAACCCCTAGCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGGCG 3582
          ||| | ||  |||||||||||||||||||||||||||||||||||||||||||||||
Db    853 GGGCTCAACGGCAGATCTCATCCTCATGGACCC-AGTCCCAGAAGCTGTACTGGCG 909

RESULT 3
BM804831                                1004 bp    mRNA    linear   EST 05-MAR-2002
LOCUS       BM804831
DEFINITION  AGENCOURT_6484845 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725541
            5', mRNA sequence.
ACCESSION   BM804831
VERSION     BM804831.1  GI:19121654
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation

            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: MAM12716  row: j  column: 06
            High quality sequence stop: 625

FEATURES             Location/Qualifiers
     source          1..1004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5725541"
                     /clone_lib="NIH_MGC_125"
                     /lab_host="DH10B"
                     /note="Organ, ovary (pool of 3); Vector: pCMV SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source: pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."

BASE COUNT      231 a    347 c    269 g    157 t
ORIGIN

  Query Match          18.1%;  Score 754;  DB 14;  Length 1004;
  Best Local Similarity   96.6%;  Pred. No. 2.2e-97;
  Matches  813;  Conservative    0;  Mismatches   26;  Indels    4;  Gaps    4;

Qy    382 GAGGAGCAAGAAGTGGAAACAAGAGCAGCGGCCGTCCCCTCCCCCCCATCGTGAAG 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGGAGCAAGAAGTGGAAACAAGAGCAGT[illegible]ATCGTGAAG 60

Qy    382 AAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAG 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAG 120

Qy    442 GAGACAGTGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCTGGGGAGCTGTGGTGCGC 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGACAGTGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCTGGGGAGCTGTGGTGCGC 180

Qy    502 AG-TCAGACTCCAAGGCCCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCGGC 560
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCTCAGACTCCAAGGCCCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCGGC 240

Qy    561 ATGTACCAGAGCGGCCTGGGGCTGCCTGTGCCGTGCAGCCATCGAGGCTCCCAAA 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATGTACCAGAGCGGCCTGGGGCTGCCTGTGCCGTGCAGCCATCGAGGCTCCCAAA 300

Qy    621 GCTTTCCTGAGGAAAATCGACGCCAAGGACGAGGCTCTGGCCAAGGTCCTCATCAAGGGT 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCTTTCCTGAGGAAAATCGACGCCAAGGACGAGGCTCTGGCCAAGGTCCTCATCAAGGGT 360

Qy    681 GCCCACTCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCCCGCCAGCTGTG 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCCCACTCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCCCGCCAGCTGTG 420

Qy    741 GCTCCTCGACCCAAGGCCCCGCTACAGCTTGGGCCCTCTAGCTCCATCAAGGAAAAGCAG 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCTCCTCGACCCAAGGCCCCGCTACAGCTTGGGCCCTCTAGCTCCATCAAGGAAAAGCAG 480

Qy    801 GGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACAGACCCCCACCTCCA 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACAGACCCCCACCTCCA 540

Qy    861 CCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGC 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGC 600

Qy    921 TTGACACAGCCCGTGGAGGACCAGGGGTCTCCACCCAGCTACTGGCGCCCTCTGGCAGC 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTGACACAGCCCGTGGAGGACCAGGGGTCTCCACCCAGCTACTGGCGCCCTCTGGCAGC 660

Qy    981 GTGTGCTTCTCCTACACCGGCAGG-CCCTGGAAGTTGTTCCTACGCAAGGAGCTGTTCTA 1039
```

```
RESULT 4
BI773824
LOCUS       BI773824                  799 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  yx95f10.s1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5205535 5',
            mRNA sequence.
ACCESSION   BI773824
VERSION     BI773824.1  GI:15763902
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11b15  row: g  column: 08
            High quality sequence start: 5
            High quality sequence stop: 799
FEATURES             Location/Qualifiers
     source          1..799
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5205535"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source:
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20 22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning).  Average insert size 1.4 kb, insert size
                     range 1-2 kb.  Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      138 a    247 c    240 g    174 t
ORIGIN
```

```
  Query Match:          17.6%;  Score 736.6;  DB 11;  Length 799;
  Best Local Similarity:   98.4%;  Pred. No. 0.0e+0;
  Matches  734;  Conservative    3;  Mismatches    9;  Indels    4;  Gaps    4;

Qy    1915 CATGTGTTAAGACTACCCTGCTGTGCCCTCCTCCTGCCCCTCCCCTGCCTCCTGAT 1974
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5 CATGTGGTGGGA-TACCCTGCTGTGCCGTCCTCCTGTGGCTG-CCCTCTGCCCGCCTCCTGAT 62

Qy    1975 GAGTGCCTTGTGTCCTCACCAACACTCTGTCAATCCTTGCAGGCAGGGGCTGGGCTGGAT 2034
```

```
Db      63 GGCTGGCTTGTCTCCACAAGAGAGTGTGACTCTGTCCACCTCTGCCTCCTGAT 117

Qy    2035 GCTGCTCTTGTGTCCTACGTGGTACTTAGTCAAAGATCCTTAGTAAATAAA 2084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 GCTGCTCTTGTGTCCTACGTGGTACTTAGTCAAAGATCCTTAGTAAATAAA 167

Qy    2085 CATGTCTTCACTTTCCTGGCTTCTGGTCTTCTTTGTTTTTTTTTTTTGGGAGGGATC 2144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     168 CATGTCTTCACTTTCCTGGCTTCTGGTCTTCTTTGTTTTTTTTTTTTGGGAGGGATC 227

Qy    2145 GGGCTGGGGCAGGACCCCGGGACAGGGCACTGGACACTCATAGGGGCACAGGTTTTTC 2204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     228 GGGCTGGGGCAGGACCCCGGGACAGGGCACTGGACACTCATAGGGGCACAGGTTTTTC 287

Qy    2205 TGATCCCAGGGCCCTGCCCACCCTTGGAGCCAGGTACACACTCATCTGGAGGCCACCC 2264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     288 TGATCCCAGGGCCCTGCCCACCCTTGGAGCCAGGTACACACTCATCTGGAGGCCACCC 347

Qy    2265 ACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAAA 2324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     348 AGCCTGTCCTCCTGTGGGCCTATGCCTTTCTGCCCTGACTCCTACACTCATCTCTAAA 407

Qy    2325 TTCGCCCGGGCGTTACTTCCGGAAGGTCCCAGGCCTTGCTGGGCCAGACTGATGGAGGTCC 2384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     408 TTCGCCCGGGCGTTACTTCCGGAAGGTCCCAGGCCTTGCTGGGCCAGACTGATGGAGGTCC 467

Qy    2385 GCAGGAAAGGACAGGAGTAGCTTGGGTAGTACACAAGGTCTCTATCTAGAGTGGCTC 2444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     468 GCAGGAAAGGACAGGAGTAGCCTGGGTAGTACACAAGGTCTCTATCTAGAGTGGCTC 527

Qy    2445 CTCAGCCTCAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGTTTCTTGTCTATCTCA 2504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     528 CTCAGCCTCAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGTTTCTTGTCTATCTCA 587

Qy    2505 GTTTATGGGTGACCAGTCCAAGCCCCGGGGCAAGGATGAGATGATCTCTATGAATT 2564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     588 GTTTATGGGTGACCAGTCCAAGCCCCGGGGCAAGGATGAGATGATCTCTATGAATT 647

Qy    2565 GCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGATGAGATTTATTTTAATTTATGAATA 2624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     648 GCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGATGAGATTTATTTTAATTTATGAATA 707

Qy    2625 TGTGTACAATACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     708 GGTCACGGGACACCCCCGG CGGAACACTGTACTCGAGGCTGGAAAA TTGTTGTAGTTTGT 767

Qy    2685 CACAAGGCTTCTTCCCCCCG 2713
           |||||||||||||||||||||
Db     768 CACAGGCTTCTTCCCCCCG 799
```

```
RESULT 5
BI766913
LOCUS       BI766913                  1070 bp    mRNA    linear   EST 27-SEP-2001
DEFINITION  yx95f10.r1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5205535 5',
            mRNA sequence.
ACCESSION   BI766913
VERSION     BI766913.1  GI:15758491
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1070)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM11508  row: o  column: 15
          High quality sequence start: 2
          High quality sequence stop: 817.
FEATURES             Location/Qualifiers
     source          1. .1070
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5203046"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20 22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking code 023. Note:
                     this is a NIH_MGC Library."

BASE COUNT      216 a    361 c    285 g    208 t
ORIGIN

  Query Match             17.6%;  Score 735.8;  DB 13,  Length 1070,
  Best Local Similarity   90.7%;  Pred. No. 8e-95;
  Matches  908;  Conservative    0;  Mismatches   77;  Indels   16,  Gaps   11;

Qy     567 CAGAGCCGCCCCGGGCCCCGTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAAGCTTTC 626
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAGAGCCGCCCCGGGCCCCGTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAAGCTTTC 60

Qy     627 CTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGCTGGGTATCAACGGTGCCCAC 686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGCTGGGTATCAACGGTGCCCAC 120

Qy     687 TCGTCCCCGCCGATGCTGTCCCCAGCCCAGGAAAGGGCCCCCCGCCAGCTGTGGCTCCT 746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCGTCCCCGCCGATGCTGTCCCCAGCCCAGGAAAGGGCCCCCCGCCAGCTGTGGCTCCT 180

Qy     747 CGACCCAAGGCCCCCTACAGCTTGGGCCCTCTAGCTCCATCAAGGAAAAGCAGGGGCCC 806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGACCCAAGGCCCCCTACAGCTTGGGCCCTCTAGCTCCATCAAGGAAAAGCAGGGGCCC 240

Qy     807 CTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACACACCCCCACCTCCACCAGCG 866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACACACCCCCACCTCCACCAGCG 300

Qy     867 CCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGCTTGACA 926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGCTTGACA 360

Qy     927 CAGCCCGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGCGCCCTCTGGCAGCGTGTGC 986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAGCCCGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGCGCCCTCTGGCAGCGTGTGC 420

Qy     987 TTCTCCTACACCGGCACCCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGG 1046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTCTCCTACACCGGCACCCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGG 480

Qy    1047 GAGAACTTCAGCCATGCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACC 1106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAGAACTTCAGCCATGCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACC 540

Qy    1107 TTCTCCGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTG 1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTCTCCGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTG 600

Qy    1167 GGA--GGCTTGGAGGTGGACCTGGATTCTCTCACCACCACCGAAGACAG-CGTCAAGAAG 1223
           |||  | ||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     601 GGAGCGCTTTGGAGGTGGACCTGGATTCTCTCACCACCACCGAAGACAGCCGTCAAGAAG 660

Qy    1224 CGCATCGTGGTGGCCGCTC-GGGACAACTGGGCCAATTA--CTTCTCCCGCTTCTTTCC- 1279
           |||||||||||||||||| ||||||||||||||||||||    ||||||||||||||||| 
Db     661 CGCATCGTGGTGGCCGCTCGGGGACAACTGGGCCAATTAACTTTCTCCCGCTTCTTTCCT 720

Qy    1280 TGTCTCGGGCGAGAGTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCG 1339
           |||||||||||||||||||| |||||||||| ||||||||| |||||||||||||||| 
Db     721 TGTCTCGGGCGAGAGTGGCAGC-ACGTGCAGCTGTAAGCCGTGT CCACCGTGGGCTGCG 778

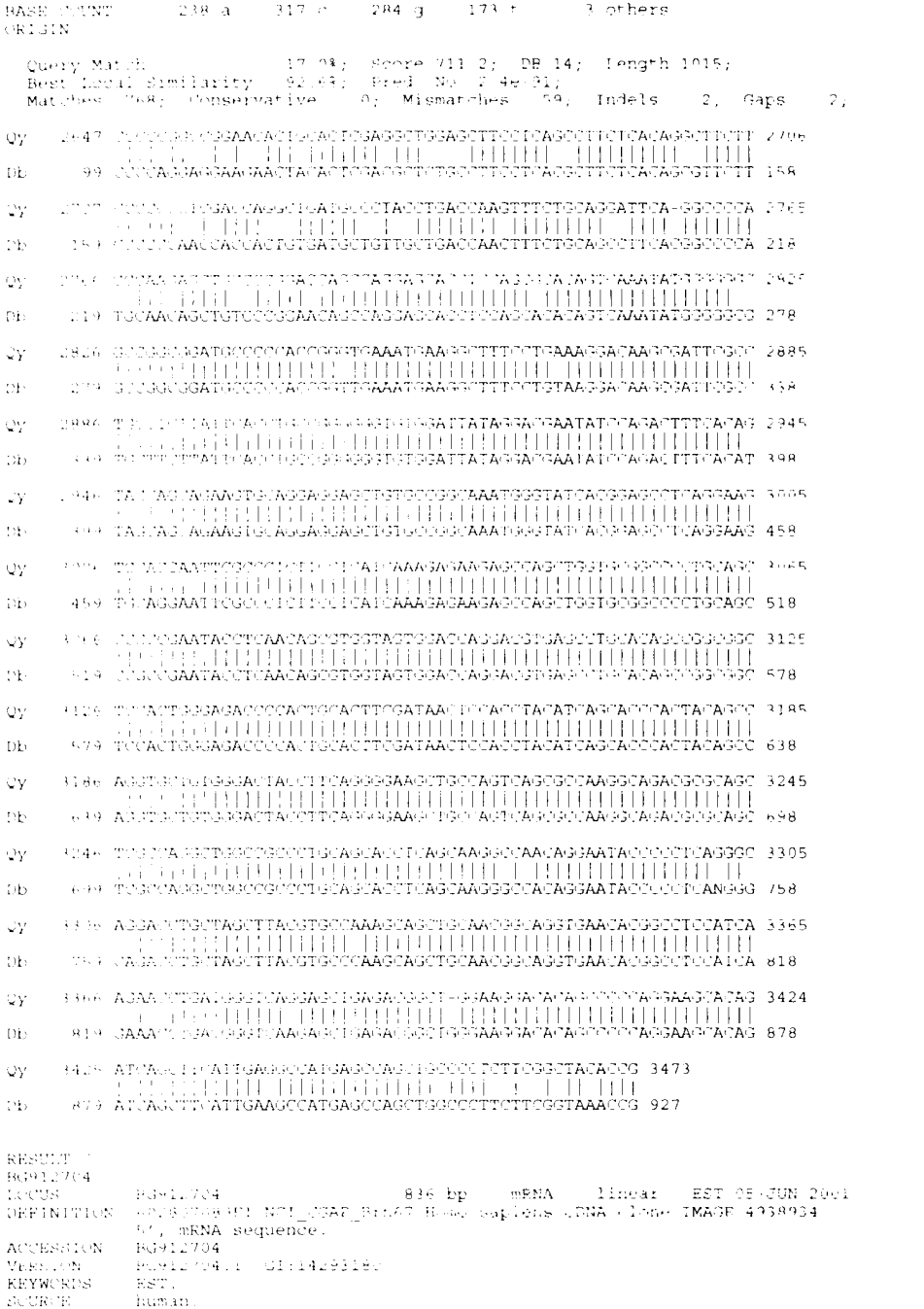
Qy    1340 ACTGCTCAAGGTGACCCAAGGCCCCGGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTC 1399
           |||   | |   |||||| |||||   || |||| |||||||||||||||| | ||| 
Db     779 ACTTGCTCAAGGTGACCAAGGACCCGGC--TCGGCCCGAACAGCTGAAGATTCTC-GGTC 835

Qy    1400 ATACAGCTTTGCGGAGGTGCTGGGTGTGGAGTGCCGGGGCCGCTCCACCCTGGAGCTGTC 1459
           ||||| || |||   |    ||  | ||||||  || |||  |||||||||||||||| 
Db     836 ATACAGTTTGGCGAGTG---CTGGGTTTGAGTGCGA[illegible]GTCCCACCCTGGAGCTGTC 892

Qy    1460 ACTGAAGAG-CGAGCAGCTGGTGCTGCACACAG[illegible]GAAGGCGATCGAGGCGCTGG 1518
           || | || ||| |  | ||||   || |||||||  ||||| | | |  | | 
Db     893 CCTTGATAGCCGATACTTTGTTACTGCCACAAGGCCGGCA[illegible]AAGAAGGCGC 952

Qy    1519 TTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTAT 1559
           | | |   |  ||  ||||| |||||||  |||  | |
Db     953 GTTACGCAAATCCTAACTAGCTTTAGAAGGATCCCG[illegible]ATGT 993


RESULT 6
BM923763
LOCUS       BM923763                1015 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6709671 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5759880
            5', mRNA sequence.
ACCESSION   BM923763
VERSION     BM923763.1  GI:19374142
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1015)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12806  row: a  column: 01
            High quality sequence start: 222
            High quality sequence stop: 720.
FEATURES             Location/Qualifiers
     source          1. .1015
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5759880"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 4 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 42 yo male and 70 yo female.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
```


BASE COUNT 238 a 317 c 284 g 173 t 3 others
ORIGIN
Query Match 17.9%; Score 711.2; DB 14; Length 1015;
Best Local Similarity 92.8%; Pred. No. 2.4e-91;
Matches 768; Conservative 0; Mismatches 59; Indels 2; Gaps 2;
RESULT 1
BG912704
LOCUS BG912704 836 bp mRNA linear EST 05-JUN 2001
DEFINITION BG912704 NCI_CGAP_Bm67 Homo sapiens cDNA clone IMAGE 4938934
5', mRNA sequence.
ACCESSION BG912704
VERSION BG912704.1 GI:14299186
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 836)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10875 row: b column: 23
High quality sequence stop: 827.
FEATURES location/Qualifiers
source 1..836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4938934"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q
loss"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primers: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
BASE COUNT 185 a 247 c 249 g 155 t
ORIGIN
Query Match 16.9%; Score 703.6; DB 14; Length 836;
Best Local Similarity 90.9%; Pred. No. 8.1e-90;
Matches 821; Conservative 0; Mismatches 74; Indels 69; Gaps 4;

```
Qy   2808 CAGTCAAATATGGGGGGGGGGGGGGATGGCCCCACCGGGTGAAATGAAGGCTTTCCTGA 2867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 CAGTCAAATATGGGGGGGGGGGGGGATGGCCCCACCGGGTGAAATGAAGGCTTTCCTGA 533

Qy   2868 AAGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGA 2927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 AAGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGA 593

Qy   2928 ATATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTA 2987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    594 ATATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTA 653

Qy   2988 TCAGGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCCAGC 3047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    654 TCAGGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCCAGC 713

Qy   3048 TGGTGGGGGGGGGTGTAGCGGGCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGA 3107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    714 TGGTGGGGGGGGGTGTAGCGGGCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGA 773

Qy   3108 GCCTGCACTCCACCTAC 3124
          |||||||||||||||||
Db    774 GCCTGCACTCCACCTAC 790

Qy   3125 ATCA 3128
          ||||
Db    791 ATTA 794
```

```
RESULT 8
BI766411                                908 bp    mRNA    linear   EST 25-SEP-2001
LOCUS       BI766411                    908 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  602052657T1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201966 5',
            mRNA sequence
ACCESSION   BI766411
VERSION     BI766411.1  GI:15757984
KEYWORDS    EST
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11506  row: b  column: 15
            High quality sequence stop: 852
FEATURES             Location/Qualifiers
     source          1..908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201966"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen, Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source:
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      209 a    287 c    272 g    140 t
ORIGIN

  Query Match           16.8%;  Score 699.6;  DB 13;  Length 908;
  Best Local Similarity   91.3%;  Pred. No. 1.1e-89;
  Matches  829;  Conservative    0;  Mismatches   24;  Indels   55;  Gaps    6;

Qy   2629 CAAGCAGGTCACAGGACACCCCCGGCCGGAACACTGCACTCGA---TGGAGCTTCCTCAG 2685
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAAGCAGGTCACGGGACACCCCCGGCCGGAACACTGCACTCGAGGCTGGAGCTTCCTCAG 60

Qy   2686 CCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGATGCCCTACCTGACCAAGTTTCT 2745
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGATGCCCTACCTGACCAAGTTTCT 120

Qy   2746 GCAGGATTCAGGCCCCAGCCAAGAAGCTGGCCGAGATCATCAACCAGGCCAC 2795
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCAGGATTCAGGCCCCAGCCAAGAAGCTGGCCGAGATCATCAACCAGGCCAC 170

Qy   2796 AGTCAAATATGGGGGGCGCCGGCGGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAA 2855
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 AGTCAAATATGGGGGGCGCCGGCGGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAA 230

Qy   2856 AAGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGAA 2915
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 AAGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGAA 290

Qy   2916 TATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTAT 2975
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTAT 350

Qy   2976 CACGGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGA------ 3029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||      
Db    351 CACGGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGTAAGCT 410

Qy   3030 ------------------------------------  --  GCCAGCTGGTGGCGGCCCC 3059
                                                    ||||||||||||||||||||
Db    411 CACGGAGCCTCAGGAAGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGTAAGCT 470

Qy   3060 TGCAGCCCGCCGGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCC 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 TGCAGCCCGCCGGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCC 530

Qy   3120 GGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCAGCAACCCACT 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 GGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCAGCAACCCACT 590

Qy   3180 ACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAGCTGCCAGTCAGCGGCCAAGGCAGACG 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 ACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAGCTGCCAGTCAGCGGCCAAGGCAGACG 650

Qy   3240 CGCAGCTCGCCAGGCTGGCCGCCTTCAGCACCTCAGCAAGGCCAACAGGAATACCCCCT 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 CGCAGCTCGCCAGGCTGGCCGCCTTCAGCACCTCAGCAAGGCCAACAGGAATACCCCCT 710

Qy   3300 CAGGGCAGGACCTGCTAGCTTACGTGCCAAAGCAGCT-GCAACGGCAGGTGAACACGGCC 3358
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    711 CAGGGCAGGACCTGCTAGCTTACGTGCCAAAGCAGCTGGAAACGGCAGGTGAACACGGCC 770

Qy   3359 TCCATCAAGAACCTGATGGG--TCAGGAGCTGAGACGGCTGGAA-GGACACAGCCCCCAG 3415
          |||||||||||||||||||  |||||||||| |||||||||| ||||||||||||||||
Db    771 TCCATCAAGAACCTGATGGGGTCAGGGAGCTGAGAAGGCTGGAAGGGACACAGCCCCCAG 830

Qy   3416 GAAGCACAGAT-CAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGT 3474
          ||||||||||| |||||||||||||||||||||||||||||| ||| | | | | | |||
Db    831 GAAGCACAGATCCAGCTTCATTGAGGCCATGAGCCAGCTGCCCATCTACGGGTAAACGG 890

Qy   3475 CTATGGGG 3482
          ||||||||
Db    891 CTATGGGG 898
```

```
RESULT 9
BQ574071/c
LOCUS       BQ574071                715 bp    mRNA    linear   EST 10 JUN 2002
DEFINITION  UI H-EZ0-bay-a-10 0-UI.s1 NCI_CGAP_Ch1 Homo sapiens cDNA clone
            UI H-EZ0 bay a-10 0 UI 3', mRNA sequence.
ACCESSION   BQ574071
VERSION     BQ574071.1  GI:21477299
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
            Orthoapedics
             cDNA Librar, preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 1-21, >AT_rich#Low_complexity (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..715
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI H EZ0 bay a 10 0 UI"
                     /clone_lib="NCI_CGAP_Ch1"
                     /tissue_type="Chondrosarcoma Grade II"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker, Site_1: EcoR I, Site_2: Not I,
                     NCI_CGAP_Ch1 is a cDNA library containing the following
                     tissue(s): Chondrosarcoma Grade II. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3 Pac vector. The oligonucleotide used to prime
                     the synthesis of first strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_LIB=UI-H-EZ0
                     TAG_TISSUE=grade-2-chondrosarcoma
                     TAG_SEQ-ATCTAATATC"
BASE COUNT      140 a    226 c    207 g    141 t      1 others
ORIGIN

  Query Match             16.7%;  Score 696;  DB 14;  Length 715;
  Best Local Similarity   39.3%;  Pred. No. 3.9e-89;
  Matches  696;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1412 GGAGTTGCTGGTGTGGAGTGGGGGGGG TCCACCCTGGAGCTGTCACTGAAGAGCGA 1471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 GGAGTTGCTGGTGTGGAGTGGGGGGGGTTCCACCCTGGAGCTGTCACTGAAGAGCGA 656

Qy    1472 GAAGGTGGTGCTGCACAGCCCGGGCAAGGGCCATGAAGGGCTGGTTGAGCTATTCCT 1531
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 GCAGCTGGTGCTGCACAGCCCGGGCAAGGGCCATGAAGGGCTGGTTGAGCTATTCCT 596

Qy    1532 GAATGAGCTTAAGAAGGACTCGGGCTATGTCATCGCCCTGGCAGCTACATCACTCACAA 1591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 GAATGAGCTTAAGAAGGACTCGGGCTATGTCATCGCCCTGGCAGCTACATCACTCACAA 536

Qy    1592 CTGCAGGCTGCCTCAGCTTCCAGCCTGGGGACCTCATCAAGTCTGCCGGTGGCACCCT 1651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CTGCAGGCTGCCTCAGCTTCCAGCCTGGGGACCTCATCAAGTCTGCCGGTGGCACCCT 476

Qy    1652 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCCCATCTTTCCTTCATCAT 1711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCCCATCTTTCCTTCATCAT 416

Qy    1712 ACTCCAGCCCCCCTCCCCCCCCCCACTTTCCTCCTCCCAAACACCAGCAGCAGCTGGCTGCCA 1771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ACTGCAGCCCGCCTCCCCACTTTCCTCCTCCAAACACCAGCAGCAGCTGGCTGCCA 356

Qy    1772 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGGGCAGCAGGCCTCAGAGT 1831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGGGCAGCAGGCCTCAGAGT 296

Qy    1832 GAGGAAGATGGGAGAGGACAAGAAGCAAGGCTCTCCTGAACTGAATAAAGAAAGT 1891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 GAGGAAGATGGGAGAGGACAAGAAGCAAGGCTCTCCTGAACTGAATAAAGAAAGT 236

Qy    1892 GGTTTGATCATTTTAATCTTTCATCTGGTGGAACTCTCTCTTTTTTTTTTTTTTTTTT 1951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGTTTGATCATTTTAATCTTTCATCTGGTGGAACTCTCTCTTTTTTTTTTTTTTTTTT 176

Qy    1952 CTTGCCCCTCTGCCCGCTCCTGATGGCTCGCTTTGTCTCTCCAGCAAGAGCTTAAGTCA 2011
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 CTGCCCCTCTGCCCGCTCCTGATGGCTCGCTTTGTCTCTCCAGCAAGAGCTTAAGTCA 116

Qy    2012 TGCAGGCAGGGCTGGGCTGGATGCTGCTCTTGTGTCCAGCCTTAAGTCAAGGCTTAAT 2071
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 TGCAGGCAGGGCTGGGCTGGATGCTGCTCTTGTGTCCAGCCTTAAGTCAAGGCTTAAT 56

Qy    2072 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 2108
           ||||||||||||||||||||||||||||||||||||||
Db      55 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 19


RESULT 10
BM977966/c
LOCUS       BM977966                689 bp    mRNA    linear   EST 31 MAR 2003
DEFINITION  UI CF EC1-aec d 23 0 UI s1 UI CF EC1 Homo sapiens cDNA clone
            UI CF EC1-aec d 23 0 UI 3', mRNA sequence.
ACCESSION   BM977966
VERSION     BM977966.1  GI:19536816
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2224 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
```

```
  source          1. .689
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="UI-CF-EC1-aec-d-23-0-UI"
                  /clone_lib="UI-CF-EC1"
                  /tissue_type="Lung"
                  /dev_stage="Adult and Fetal"
                  /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                  /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                  modified polylinker; Site_1: EcoR I; Site_2: Not I,
                  UI-CF-EC1 is a normalized cDNA library containing the
                  following tissue(s): Normal lung from adult and from fetal
                  day 64, day 87, week 19 and week 42. The library was
                  constructed according to Bonaldo, Lennon and Soares,
                  Genome Research, 6:791-806, 1996. First strand cDNA
                  synthesis was primed with an oligo-dT primer containing a
                  Not I site. Double stranded cDNA was ligated to an EcoR I
                  adaptor, digested with Not I, and cloned directionally
                  into pT7T3-Pac vector. The oligonucleotide used to prime
                  the synthesis of first-strand cDNA contains a library tag
                  sequence that is located between the Not I site and the
                  (dT)18 tail. The sequence tag for this library is
                  AAGTGCTTAC.
                  TAG_LIB=UI-CF-EC1
                  TAG_TISSUE.Normal Lung Epithelial Cells Tissue nos 369-371
                  and 380-383
                  TAG_SEQ=AAGTGCTTAC"
BASE COUNT      105 a    192 c    225 g    166 t      1 others
ORIGIN

  Query Match             16.4%;  Score 683.2;  DB 14;  Length 689;
  Best Local Similarity   99.4%;  Pred. No. 2.5e-87;
  Matches  685;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    3479 GGGGTGTTGCGAGTGAGCATGCAGGGCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGC 3538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 GGGGTGTTGCGAGTGAGCATGCAGGGCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGC 630

Qy    3539 CAGCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAG 3598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 CAGCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAG 570

Qy    3599 AGCCTGCAGGGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGGCCTGGAA 3658
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     569 AGCCTGCAGGGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGGCCTGGAA 510

Qy    3659 GTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGCTGCCACAGGCCCAG 3718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 GTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGCTGCCACAGGCCCAG 450

Qy    3719 GAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACTGAGTGGCCC 3778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 GAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACTGAGTGGCCC 390

Qy    3779 AGAAGAAACTTAGAGGAGTGCAGGCGGGGAGAGAAGAGGATGAGGCCTCCCCCGGCCCA 3838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 AGCATCAACTGAGAGGAGTGCAGGCGGGGAGAGAAGAGGATGAGGCCTCCCCCGGCCCA 330

Qy    3839 AATCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCA 3898
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 AGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCA 270

Qy    3899 GCACAGCCAAGCCGCCCAGATGCAGGCCATGAGGCAGGGGCTGCTATCACGTCACCAGC 3958
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 GCACAGCCAAGCCGCCCAGATGCAGGCCATGAGGCAGGGGCTGCTATCACGTCACCAGC 210

Qy    3959 AGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGA 4018
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 AGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGA 150

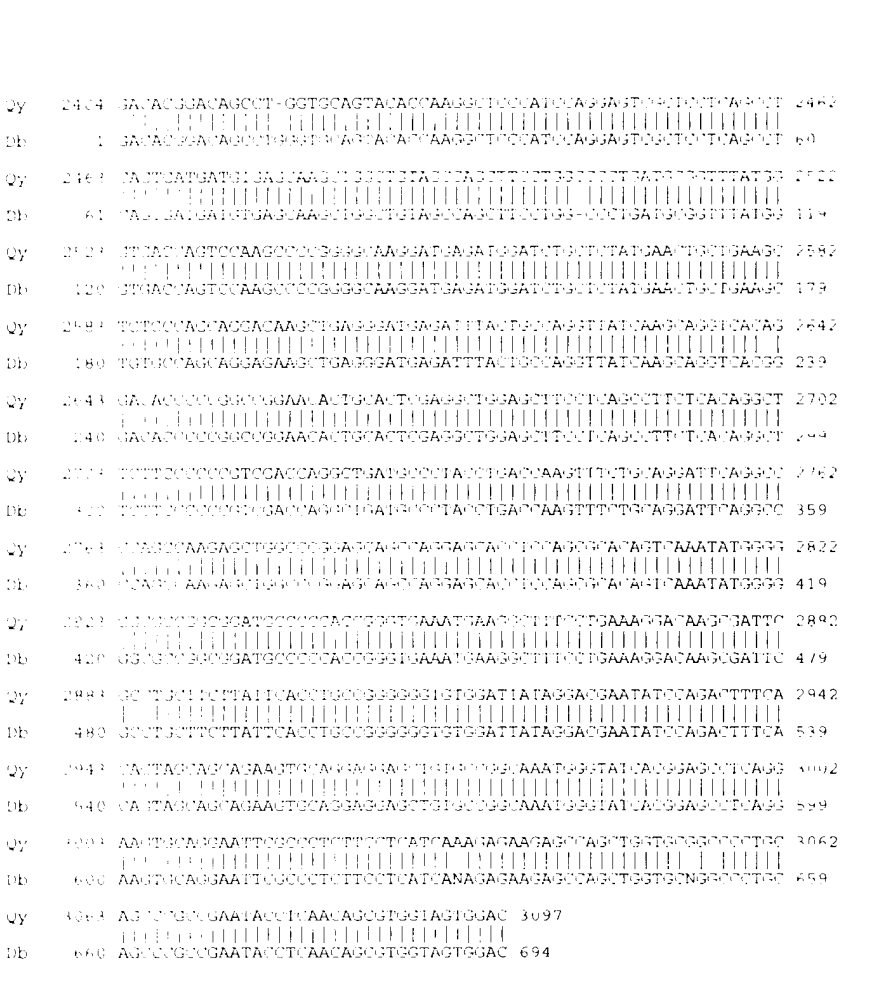
Qy    4019 ACTCGGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGACCCCAGGCT 4078
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149 ACTCGGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGACCCCAGGCT 90

Qy    4079 CCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATA 4138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 CCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATA 30

Qy    4139 AAACTCCCTAAAAAAAAAAAAAAAAAAAA 4167
           |||||||||  |||||||||||||||||
Db      29 AAACTCCCTGGAGAAAAAAAAAAAAAAAA 1
```

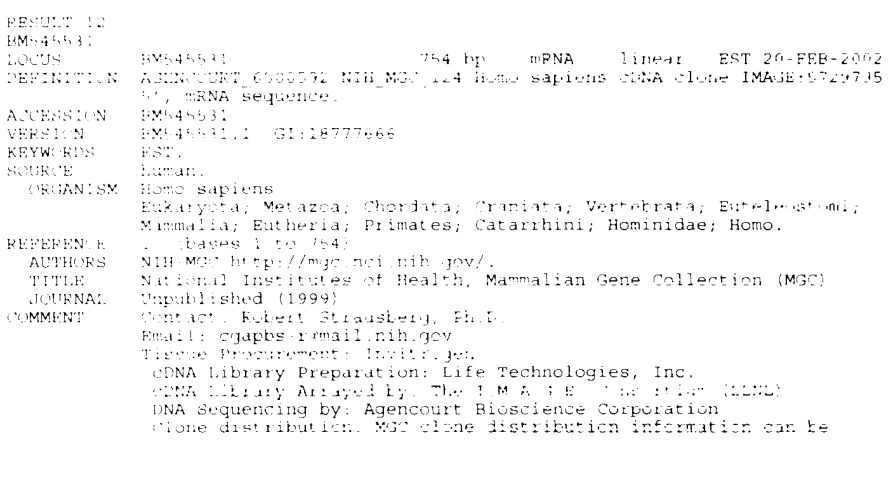
```
RESULT 11
BM720779
LOCUS       BM720779                 728 bp    mRNA    linear   EST 01-MAR-2002
DEFINITION  UI-E-EO1-aib-g-07-0-UI.r1 UI-E-EO1 Homo sapiens cDNA clone
            UI-E-EO1-aib-g 07-0-UI 5', mRNA sequence.
ACCESSION   BM720779
VERSION     BM720779.1  GI:19040260
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel. 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue weeg.uiowa.edu
            Tissue Procurement: Dr  Gregg Hageman
            cDNA Library preparation: Dr  M  Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EO1 aib-g-07-0-UI"
                     /clone_lib="UI-E-EO1"
                     /tissue_type="fetal eye"
                     /dev_stage="fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EO1 is a normalized cDNA library containing the
                     following tissue(s): fetal eye. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research. 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     CGGGTATACC. This library was created for the program, Gene
                     Discovery in the Visual System, supported by the National Eye
                     Institute (NEI)."
BASE COUNT      162 a    210 c    222 g    132 t      2 others
ORIGIN

  Query Match             16.3%;  Score 665.8;  DB 14;  Length 728;
  Best Local Similarity   99.1%;  Pred. No. 7.1e-85;
  Matches  689;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;
```



found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12727 row: k column: 12
High quality sequence start: 15
High quality sequence stop: 647.

```
FEATURES             Location/Qualifiers
     source          1..754
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5729795"
                     /clone_lib="NIH_MGC_124"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: Not I
                     (destroyed); Site 2: Not I; RNA source: male hippocampus,
                     age 27. Library is oligo dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                     normalized and enriched for full length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 012."
BASE COUNT      130 a    212 c    219 g    140 t      4 others
ORIGIN
```



```
RESULT 12
BM546531
LOCUS       BM546531                 754 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6000592 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5729795
            5', mRNA sequence.
ACCESSION   BM546531
VERSION     BM546531.1  GI:18777666
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
```

```
b     688 CTTTC 692


RESULT 13
BM971659/c
LOCUS       BM971659                  679 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-EC1-abn-i-09-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
            UI-CF-EC1-abn-i-09-0-UI 3', mRNA sequence.
ACCESSION   BM971659
VERSION     BM971659.1  GI:19589746
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..679
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EC1-abn-i-09-0-UI"
                     /clone_lib="UI-CF-EC1"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EC1 is a normalized cDNA library containing the
                     following tissue(s). Normal lung from adult and from fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site.  Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC.
                     TAG_LIB=UI-CF-EC1
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_SEQ=AAGTGCTTAC"
BASE COUNT      133 a    208 c    200 g    138 t
ORIGIN

  Query Match           15.8%;  Score 659.4;  DB 14;  Length 679;
  Best Local Similarity   99.8%;  Pred. No. 5.9e-84;
  Matches  660;  Conservative    0;  Mismatches    1;  Indels    0,  Gaps    0,

Qy    1449 CTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTGCACACAGCCCGGGCAAGGGCCATC 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     679 CTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTGCACACAGCCCGGGCAAGGGCCATC 620

Qy    1509 GAGGCGCTGGTTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGCC 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     619 GAGGCGCTGGTTGAGCTATTCCTGAACGAGCTTAAGAAGGACTCCGGCTATGTCATCGCC 560

Qy    1569 CTGCGCAGCTACATCACTGACAACTGCAGCCCTCCTCAGCTTCCACCCTGGGGACCTCATC 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     559 CTGCGCAGCTACATCACTGACAACTGCAGCCCTCCTCAGCTTCCACCCTGGGGACCTCATC 500

Qy    1629 AAGCTGCTGCCGGTGGCCACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGCCGT 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 AAGCTGCTGCCGGTGGCCACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGCCGT 440

Qy    1689 TCCGGACTCTTTCCTGCCGACATAGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCC 1748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 TCCGGACTCTTTCCTGCCGACATAGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCC 380

Qy    1749 AAGGAGCAGAGGAGTGGCTGGCACAAGGGTCAGCTGTCCAACGGGAAACCAGGGCTGGCT 1808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 AAGGAGCAGAGGAGTGGCTGGCACAAGGGTCAGCTGTCCAACGGGAAACCAGGGCTGGCT 320

Qy    1809 CGGTGGGACAGGGCCTCAGAGGTGAGGAAGATGGGAGAGGGACAAGCAGAGCCAAGGCCT 1868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 CGGTGGGACAGGGCCTCAGAGGTGAGGAAGATGGGAGAGGGACAAGCAGAGCCAAGGCCT 260

Qy    1869 GCCTGAGACTGAGGAAGGAAAGGGGTTTGACCACTCCCGAGGCTGCCATGCGGTGGGACC 1928
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GCCTGAGACTGAGGAAGGAAAGGGGTTTGACCACTCCCGAGGCTGCCATGCGGTGGGACC 200

Qy    1929 ACCCTGCTGTCCGTCTCCTGTGGCTGCCCCTCTGCCCGCTCCTGATGGCTGGCCTGGCT 1988
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 ACCCTGCTGTCCGTCTCCTGTGGCTGCCCCTCTGCCCGCTCCTGATGGCTGGCCTGGCT 140

Qy    1989 CTCCAGCAAGACTGTGCACTCCTTGCAGGCAGGGGCTGGGCTGGATGCTGCTCTTGTGTC 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CTCCAGCAAGACTGTGCACTCCTTGCAGGCAGGGGCTGGGCTGGATGCTGCTCTTGTGTC 80

Qy    2049 CCACGTGGTACTTAGTTCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 2108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 CCACGTGGTACTTAGTTCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 20

Qy    2109 C 2109
             |
Db      19 C 19


RESULT 14
BM978287/c
LOCUS       BM978287                  668 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-EC1-aeb-p-11-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
            UI-CF-EC1-aeb-p-11-0-UI 3', mRNA sequence.
ACCESSION   BM978287
VERSION     BM978287.1  GI:19597563
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
```

```
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EC1-aeb-p-11-0-UI"
                     /clone_lib="UI-CF-EC1"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified pol/linker, Site_1: EcoR I, Site_2: Not I;
                     UI-CF-EC1 is a normalized cDNA library containing the
                     following tissue(s): Normal lung from adult and fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3 Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC.
                     TAG_LIB=UI-CF-EC1
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_SEQ=AAGTGCTTAC"
BASE COUNT      131 a    205 c    196 g    135 t      1 others
ORIGIN

  Query Match             19.5%;  Score 645.8;  DB 14;  Length 668;
  Best Local Similarity   99.8%;  Pred. No. 4.9e-82;
  Matches  649;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1445 AGAAGGAGCAGCTGGTGCTGCACACAGCCCAGGCAAAGGGCATCGGAAGGCCTGGTTGAGC 1504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     668 AGAAGGAGCAGCTGGTGCTGCACACAGCCCAGGCAAAGGGCATCGGAAGGCCTGGTTGAGC 609

Qy    1505 TATTCTGAATCAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCTGGGCAACCTACATCA 1564
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     608 TATTCTGAATCAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCTGGGCAACCTACATCA 549

Qy    1565 CTGAGAACTCCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGTCCGTGG 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     548 CTGAGAACTCCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGTCCGTGG 489

Qy    1625 TCACCCTGGAGCCAGGCTGGCAGTTTGGCTCCGGGGGCCGTTCCGGACTCTTTCCTG 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     488 TCACCCTGGAGCCAGGCTGGCAGTTTGGCTCCGGGGGCCGTTCCGGACTCTTTCCTG 429

Qy    1685 CCAGACATACTGCAGCCGGCCGCCGCTCCCGACTTTTCCTTCTCCAAGGAGAGAGAGGAGTG 1744
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     428 CCAGACATACTGCAGCCGGCCGCCGCTCCCGACTTTTCCTTCTCCAAGGAGAGAGAGGAGTG 369

Qy    1745 GCTGGCACAAGGGCTCAGCTGTCCAACGGGAAGCTAGGGCTCAGGCTCTTGAGCTA 1804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     368 GCTGGCACAAGGGCTCAGCTGTCCAACGGGAAGCTAGGGCTCAGGCTCTTGAGCTA 309

Qy    1805 CAAAGTTGAGAAGATGGGAGAGGGACAAGCAGAGGAAGGCTGCCTGAGATTGAGGAA 1864
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     308 CAAAGTTGAGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCTGCCTGAGATTGAGGAA 249

Qy    1865 AAAATTGAGGAGCTTAAGAAGCACTCCAAGCATCTGGAAGAGCACGTGTCCGCCTCT 1924
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 GGAAAGGGGTTTGACCACTCCCAGGCTGCCATGCGGTTGGACCAGCTGTTCCCCTCT 189

Qy    1925 CCTGTGGCTGCCCCCTCCCCCCTCCCATGCCCCCCCCAAGGACTTCCCAGTAAGACTGTT 1984
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 CCTGTGGCTGCCCCCTCCCCCCTCCCATGCCCCCCCCAAGGACTTCCCAGTAAGACTGTT 129

Qy    1985 CACTCCTTGCAGGCAGGGGCTGGGCTGGATGCTGCTTTGTGTCCAGCTGGACTTAGT 2044
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 CACTCCTTGCAGGCAGGGGCTGGGCTGGATGCTGCTTTGTGTCCAGCTGGACTTAGT 69

Qy    2045 TCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTGGC 2114
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 TCAAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTGGC 19


RESULT 15
AI671062/c
LOCUS       AI671062                649 bp    mRNA    linear   EST 10-DEC-2000
DEFINITION  qz12f07.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2288863 3',
            similar to TR:G702496 G702496 MYOSIN XV ;, mRNA sequence.
ACCESSION   AI671062
VERSION     AI671062.1  GI:4860793
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   NCI-CGAP Group; http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael F.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 3521  Std Error: 0.00
            Seq primer: -40UP from Gibco.
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2288863"
                     /clone_lib="NCI_CGAP_Ut2"
                     /tissue_type="moderately differentiated endometrial
                     adenocarcinoma, 3 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Sall;
                     Site_2: Not1; Cloned unidirectionally. Primer: Sfi-Not.
                     Average insert size 1.85 Kb. Life Technologies Catalog #:
                     11539-012"
BASE COUNT      130 a    204 c    195 g    119 t      1 others
ORIGIN

  Query Match             19.0%;  Score 625.8;  DB 9;  Length 649;
  Best Local Similarity   98.0%;  Pred. No. 3.3e-79;
  Matches  631;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    1467 AGCGAGCAGCTGGTGCTGCACACAGCCCAGGCAAAGGGCATCGGAAGGCCTGGTTGAGCTA 1526
           |||||||||||||||||||  || ||  ||||||||||||||||||||||||||||||||||
Db     649 AGCGAGCAGCTGGTGCTGCACACAGCCCAGGCAAAGGGCATCGGAAGGCCTGGTTGAGCTA 590

Qy    1527 TTCTGAATCAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCTGGGCAACCTACATCACT 1586
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     589 TTCTGAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCTGGGCAACCTACATCACT 530
```

```
Qy   1587 GACAACTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGCCGGTGGCC 1646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    529 GACAACTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGCCGGTGGCC 470

Qy   1647 ACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCC 1706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    469 ACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCC 410

Qy   1707 GACATAGTGCAGCCGGCTGTCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGC 1766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    409 GACATAGTGCAGCCGGCTGTCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGC 350

Qy   1767 TGGCACAAGGGTCAGCTGTCCAACGGGGAAGCCAGGGCTGGCTCGGTGGGACAGGGCCTCA 1826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    349 TGGCACAAGGGTCAGCTGTCCAACGGGGAAGCCAGGGCTGGCTCGGTGGGACAGGGCCTCA 290

Qy   1827 GAGGTGAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGG 1886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 GAGGTGAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGG 230

Qy   1887 AAAGGGGTTTGACCAGTCCCGAGGCTGCCATGCGGTGGGACCACCCTGCTGTCCGTCTCC 1946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 AAAGGGGTTTGACCAGTCCCGAGGCTGCCATGCGGTGGGACCACCCTGCTGTCCGTCTCC 170

Qy   1947 TGTGGCTGCCCCTGTGCCCGCTCCTGATGGCTCGCCTTGTCTCTCCAGCAAGACTGTGCA 2006
          ||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db    169 TGTGGCTGCCCCTGTGCCCGCTCCTGATGGATCGCCTTGTCTCTCCAGCAAGACTGTGCA 110

Qy   2007 CTGCTTGCAGGCAGGGGCTGGGCTGGGTGCTGCTCTTGTGTCCCACGTGGTACTTAGTTC 2066
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    109 CTGCTTGCAGGCAGGGGCTGGGCTGGGTGCTGCTCTTGTGTCCCACGTGGTACTTAGTTC 50

Qy   2067 AAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTG 2112
          ||||||||||||||||||||||||||||||||||||||||||||||
Db     49 AAGGCTGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTTCCTG 4
```

Search completed: January 10, 2003, 01:55:30
Job time : 5016.49 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:10:54 , Search time 118.572 Seconds
(without alignments)
10795.718 Million cell updates/sec

Title: US-09-803-126-5
Perfect score: 4174
Sequence: 1 cgqcaqcagcaggctcgggc ..... aaaaaaaaaaaaaaaaaaaa 4174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

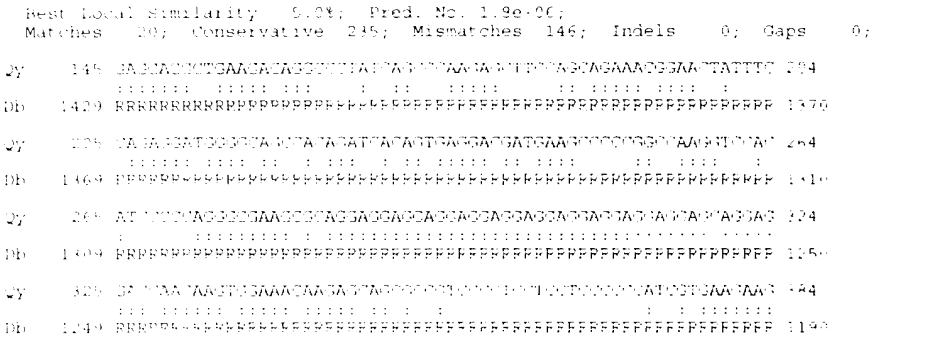
	Result No.	Score	% Query Match	Length	DB	ID	Description
C	1	73.4	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
C	2	61.2	1.5	51259	3	US-08-781-891-209	Sequence 209, App
C	3	58	1.4	16442	3	US-08-781-891-208	Sequence 208, App
	4	54.8	1.3	1362	4	US-08-979-608A-12	Sequence 12, Appl
	5	54.8	1.3	1422	4	US-08-979-608A-13	Sequence 13, Appl
	6	54.8	1.3	1617	4	US-08-979-608A-11	Sequence 11, Appl
	7	54.2	1.3	1926	4	US-09-249-585A-2	Sequence 2, Appli
	8	54.2	1.3	2580	3	US-09-050-864-2	Sequence 2, Appli
	9	54.2	1.3	2580	4	US-08-369-081-2	Sequence 2, Appli
C	10	54.2	1.3	9452	2	US-09-130-114-1	Sequence 1, Appli
	11	54.2	1.3	9600	4	US-08-910-647-1	Sequence 1, Appli
	12	54.2	1.3	9600	4	US-09-620-925-1	Sequence 1, Appli
	13	54.2	1.3	10596	1	US-07-884-811-15	Sequence 15, Appl
	14	54.2	1.3	10596	1	US-07-885-971-15	Sequence 15, Appl
	15	54.2	1.3	10596	1	US-08-087-783A-15	Sequence 15, Appl
	16	54.2	1.3	10596	1	US-08-194-088B-15	Sequence 15, Appl
	17	54.2	1.3	10596	2	US-08-194-087-15	Sequence 15, Appl
	18	54.2	1.3	10596	5	PCT-US93-04648-15	Sequence 2, Appli
	19	53.4	1.3	1276	4	US-09-177-325-2	Sequence 2, Appli
	20	53.4	1.3	1276	4	US-09-411-812A-2	Sequence 2, Appli
	21	53.4	1.3	1276	4	US-09-590-113-2	Sequence 1, Appli
	22	51.4	1.2	3489	2	US-08-728-323A-1	Sequence 1, Appli
	23	51.4	1.2	3489	4	US-09-298-568-1	Sequence 20, Appl
	24	51.4	1.2	32207	2	US-08-770-379-20	Sequence 20, Appl
C	25	51.4	1.2	32207	4	US-08-757-669A-20	Sequence 20, Appl
C	26	51.4	1.2	32207	4	US-09-230-371A-20	Sequence 23, Appl
C	27	51	1.2	2301	1	US-08-306-691B-23	Sequence 78, Appl
	28	51	1.2	2301	5	PCT-US93-06251-78	Sequence 2, Appli
	29	50.4	1.2	1678	3	US-08-650-766-2	Sequence 2, Appli
	30	50.4	1.2	1954	3	US-08-922-635-2	Sequence 4, Appli
	31	50.4	1.2	3318	3	US-08-650-766-3	Sequence 3, Appli
	32	50.4	1.2	3318	3	US-08-922-635-3	Sequence 1, Appli
	33	50.4	1.2	3385	3	US-08-650-766-1	Sequence 1, Appli
	34	50.4	1.2	3385	3	US-08-922-635-1	Sequence 21, Appl
	35	50.4	1.2	15202	3	US-08-922-635-21	Sequence 8, Appli
	36	49	1.2	19056	4	US-09-272-032-8	Sequence 53, Appl
	37	48.8	1.2	3853	4	US-09-245-248B-53	Sequence 157, App
	38	47.4	1.1	2313	4	US-09-370-838-157	Sequence 93, Appl
	39	47.4	1.1	9551	1	US-08-056-200-93	Sequence 93, Appl
	40	47.4	1.1	9551	2	US-08-800-644-93	Sequence 5, Appli
C	41	47.2	1.1	1314	4	US-09-500-569-5	Sequence 16, Appl
C	42	47	1.1	58	4	US-08-860-038-16	Sequence 16, Appl
C	43	47	1.1	58	4	US-09-580-923-16	Sequence 15, Appl
	44	46.8	1.1	58	4	US-08-860-038-15	Sequence 15, Appl
	45	46.8	1.1	58	4	US-09-580-923-15	

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
;
US-08-232-463-14

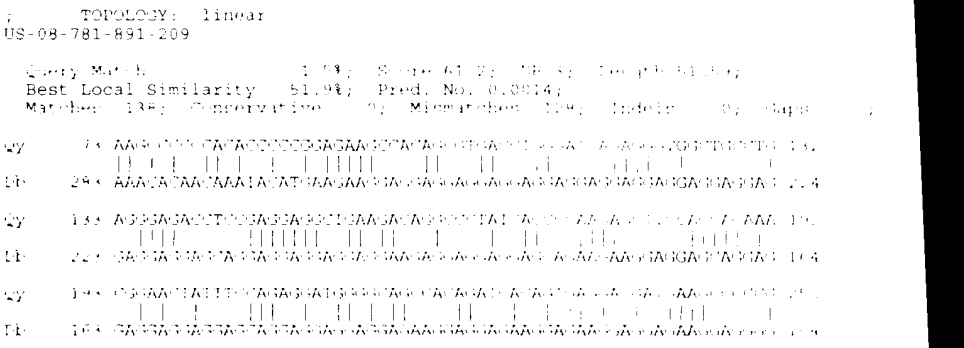
Query Match 1.8%; Score 73.4; DB 1; Length 7218;

Best Local Similarity 5.0%; Pred. No. 1.9e-06;
Matches 20; Conservative 245; Mismatches 146; Indels 0; Gaps 0;



RESULT 2
US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Ko[illegible]tenberg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240002.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-209

Query Match 1.9%; Score 61.8; DB 3; Length 91259;
Best Local Similarity 51.9%; Pred. No. 0.00014;
Matches 135; Conservative 0; Mismatches 114; Indels 0; Gaps 0;



RESULT 3
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Ko[illegible]tenberg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240002.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-208

Query Match 1.4%; Score 58; DB 3; Length 16442;

```
 Best Local Similarity   56.2%;  Pred. No. 0.0045;
 Matches  109;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     145 GAGGAGGCTGAAGACAGGCCTATCAGACACAAGAGCTTCCAGCAGAAACGGAACTATTTC 204
           ||||||| || | || | || || |||| | | | | |
Db   16411 GAGGAGGATGAGGAGGATAGGAGGAGGAGGAGGAGGAGGCAGGAGCAGGAGGAGGAG 16352

Qy     205 CAGAGGATGGGCTAGCAGATGAGATGAGTGAGAGGATGAAGCCCGGCCAAGGTCCAC 264
           ||| | || | |||  ||  | | | | | | | | | | | | | | | | | |
Db   16351 CAGGAGCAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGCAGGAGGAGGAGGAGGAG 16292

Qy     265 ATCCTCAAGGGAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAG 324
           ||| | | | |||||||||||||||||||| |||||||| |||||||||| | | |
Db   16291 CAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGCAGGAGGAGGAGGAGCAG 16238

Qy     325 GAGAAGAAGTGGA 338
           ||||| || | |||
Db   16231 GAGCAGGAGGAGGA 16218


RESULT 4
US-08-979-608A-12
; Sequence 12, Application US/08979608A
; Patent No. 6355451
;   GENERAL INFORMATION:
;        APPLICANT: Lees, Ann M.
;                   Lees, Robert S.
;                   Law, Simon W.
;                   Arjona, Anibal A.
;        TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                            BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                            TREATING ATHEROSCLEROSIS
;        NUMBER OF SEQUENCES: 42
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 225 Franklin Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110-2804
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,608A
;             FILING DATE: 26-No. 6355451-1997
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/048,547
;             FILING DATE: 03-JUN-1997
;             APPLICATION NUMBER: US 60/031,930
;             FILING DATE: 27-NOV-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Myers, Louis
;             REGISTRATION NUMBER: 35,965
;             REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 398/53818)
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617/542-5070
;             TELEFAX: 617/542-8906
;   INFORMATION FOR SEQ ID NO: 12:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1362 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;             NAME/KEY: Coding Sequence
;             LOCATION: 1..696
;        SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-08-979-608A-12

  Query Match             1.3%;  Score 54.8;  DB 4;  Length 1362;

 Best Local Similarity   48.1%;  Pred. No. 0.0096;
 Matches  155;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy     277 GAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAGGAGGAAGAAGTG 336
           |||  | |||||||||||||||||||||||| || || ||||| | || || | | |||
Db      52 GAAAAGGAGGAGGAGGAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGAGGAAGCAG 396

Qy     337 GAAACAAGAGCAGGGTGGTGGTTGTTGTTGTTGTTGTTGCAGTTGAAGAAGCCATTGAAGCAA 396
           |  | | | | | ||          || ||| |  | | |||          ||
Db     112 TCCGAGGGTTGGGAGTGGGGAGAGGGATGGTGTGGGGTTGTGAGTATTACAGCTG 171

Qy     397 GGTGGGGGCCAAAGCTCCAAAAGAGGGTGAGGTTGAGGTTGAGTTAAGTAAGTGTGT 456
           |||  | | |              || | |    |||| |||||||||    ||
Db     172 AATGGTGGGAGGGTGGGGTGTAGAGTGGTAAGGAGTGGGTAAGGAGTGGTTGTGTGT 231

Qy     457 GGCCATGGCCAAGGGCCAGGCCAAGGCAGGGATTGTGGTGTTGTTAAGC 516
           ||||    || || |  |  |||| ||||| | |  | |||||    || |
Db     232 GGCCCCCACCCTGGCCAGGAGGAAGGGTGGGGGCCGGCCGCGGGCAGTGGCACCCGCCAG 291

Qy     517 CCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGCATGTACCAGAGCCGCC 576
           | |          |  || ||   |||| || | |  |||| |  |
Db     292 GTGTTCTCCATGGCGGCCTTGAGTAAGGAGGGGGGATCAGCCTCTTCGACCACCGGGCCT 351

Qy     577 CGGGCCCCGTGCCTGTGCCCGT 598
           |||||| || |   || |
Db     352 GACTCCCCGTCCCCGGTGCCTT 373


RESULT 5
US-08-979-608A-13
; Sequence 13, Application US/08979608A
; Patent No. 6355451
;   GENERAL INFORMATION:
;        APPLICANT: Lees, Ann M.
;                   Lees, Robert S.
;                   Law, Simon W.
;                   Arjona, Anibal A.
;        TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                            BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                            TREATING ATHEROSCLEROSIS
;        NUMBER OF SEQUENCES: 42
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 225 Franklin Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110-2804
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,608A
;             FILING DATE: 26-No. 6355451-1997
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/048,547
;             FILING DATE: 03-JUN-1997
;             APPLICATION NUMBER: US 60/031,930
;             FILING DATE: 27-NOV-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Myers, Louis
;             REGISTRATION NUMBER: 35,965
;             REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 398/53818)
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617/542-5070
;             TELEFAX: 617/542-8906
;   INFORMATION FOR SEQ ID NO: 13:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1422 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
```

; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..756
; SEQUENCE DESCRIPTION: SEQ ID NO: 13;
US-08-979-608A-13

Query Match 1.3%; Score 54.8; DB 4; Length 1422,
Best Local Similarity 48.1%; Pred. No. 0.0047;
Matches 155; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 277 GAAGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGTG 336
Db 1 GAAAAGGAGGAGGAGGAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGAGGTGGTG 171

Qy 337 GAAACAAGAGCAGCGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCATTGAAGCAA 396
Db 172 TTTGAGGGTTGTGAGGTCCCCGAAGCCGATGTCCCCGGGGTGGGAGCATCACCAGCTG 231

Qy 397 GTTGGGCCAAAGCCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAG 456
Db 232 AATGGCGGGCAGCCCGGCCCGCAGACCGCCAAGGAGGGGGCAAGGAGTGGTCGTGTGT 291

Qy 457 GGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGACTTGTGGTGGCAGTCAGACTCCAAGC 516
Db 292 GGCCCTCACCCTGGCCAGGAGGAAGGGCTGGGGCCGGCCGGGGCAGTGGCACCCGCCAG 351

Qy 517 CCAAGGGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCCATGTATCAGAGCCGCC 576
Db 352 GTGTTCTCCATGGCGGCCTTGAGTAAGGAGGGGGATCAGCCTCTTCGACCACCGGCCT 411

Qy 577 CGGGCCCGTGCCTGTGCCCGT 598
Db 412 GACTCCCGTCCCCGGTGCCTT 433

RESULT 6
US-08-979-608A-11
; Sequence 11, Application US/08979608A
; Patent No. 6355461
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26 No. 6355461 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,999
; FILING DATE: 27 NOV 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,955
; REFERENCE/DOCKET NUMBER: 07[illegible]-002001 (formerly 3368/59618)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-08-979-608A-11

Query Match 1.3%; Score 54.8; DB 4; Length 1617;
Best Local Similarity 48.1%; Pred. No. 0.01;
Matches 155; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 277 GAAGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGTG 336
Db 307 GAAAAGGAGGAGGAGGAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGAGGTGGTG 171

Qy 337 GAAACAAGAGCAGCGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCATTGAAGCAA 396
Db 172 TCCGAGGGCTCGGAGGTGCCCGAGAGCGATGTCCCCGGGGTGGGAGCATCACCAGCTG 231

Qy 397 GGTGGGGCCAAAGCCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAG 456
Db 427 AATGGCGGGCAGCCGCGGCCCGCAGACCGCCAAGGAGGGGGCAAGGAGTGGTCGTGTGT 486

Qy 457 GGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGACTTGTGGTGGCAGTCAGACTCCAAGC 516
Db 487 GGCCCTCACCCTGGCCAGGAGGAAGGGCTGGGGCCGGCCGGGGCAGTGGCACCCGCCAG 546

Qy 517 CCAAGGGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCCATGTATCAGAGCCGCC 576
Db 547 GTGTTCTCCATGGCGGCCTTGAGTAAGGAGGGGGATCAGCCTCTTCGACCACCGGCCT 606

Qy 577 CGGGCCCGTGCCTGTGCCCGT 598
Db 607 GACTCCCGTCCCCGGTGCCTT 628

RESULT 7
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Hurlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF [illegible]
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA 1 DNA
US-09-249-585A-2

Query Match 1.3%; Score 54.2; DB 4; Length 1926;
Best Local Similarity 45.7%; Pred. No. 0.014;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

Qy 88 CCGGAGAAGCCACAGCGTGACCTGCCATCAGAGAGGTGGTCGCCAGAGCTCCGAG 147
Db 644 CAGGAGGGGCAGGAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 703

Qy 148 GAGGCTGAAGCAGAGGCGTATGAGGCCAAGAAGTTCCAGCAAAAAGTTCTGAG 207

```
Qy      704 CAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAG 763
Db     1027 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1086

Qy      208 AGGATGGGGCAGCCACAGATCACAGTGAGGAGATGAAGCCCCGGCCAAGGTCCACATC 267
Db     1147 CAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1206

Qy      268 CCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAG 327
Db     1207 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGG 1266

Qy      328 CAAGAAGTGGAAACAAGAGCAGCCGGTCGTCGTCGTCGTCGTCGCCATCGTGAAGAAGCCA 387
Db     1267 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1326

Qy      388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447
Db     1327 GAGGGCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 1437

Qy      448 GCGGCCAAGGGCCATGGCAAGGGCCAGGAAGGTTGGAGGCCACTGTGGTG 498
Db     1387 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 1437
```

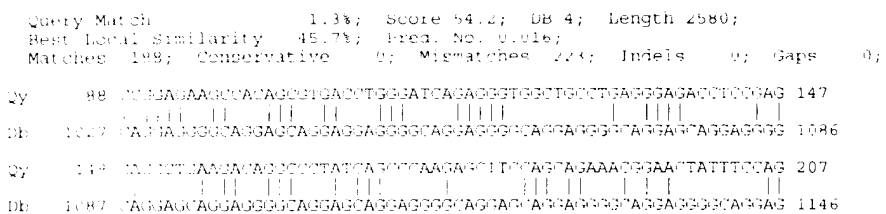
RESULT 8
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; TITLE OF INVENTION: System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 1.3%; Score 54.2; DB 3; Length 2580;
Best Local Similarity 45.7%; Pred. No. 0.016;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```
Qy       88 TCTAGAAAGTCACAGCGTCAGCTGGGATCAGAGGGTGGCTGCCTGAGGGAGACCTCCGAG 147
Db     1027 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1086
```

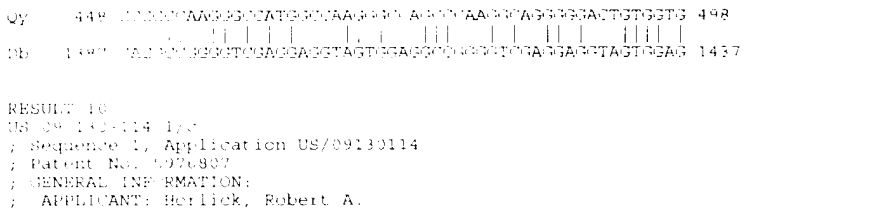
RESULT 9
US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; Hiang, Betty
; Payan, Don
; TITLE OF INVENTION. Mammalian Protein Interaction Cloning
; System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP. 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/359,081
; FILING DATE: 22-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/050,863
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

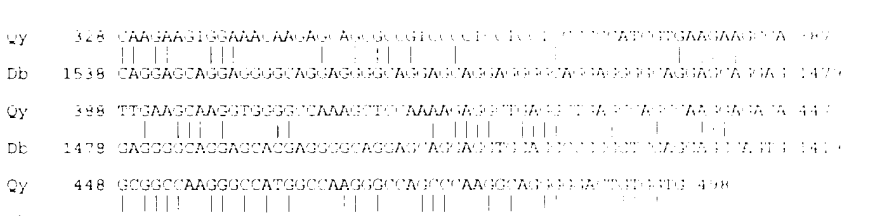
Query Match 1.3%; Score 54.2; DB 4; Length 2580;
Best Local Similarity 45.7%; Pred. No. 0.016;
Matches 198; Conservative 0; Mismatches 224; Indels 0; Gaps 0;



RESULT 10
US-09-130-114-1/5
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Herlick, Robert A.
; APPLICANT: Daouj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VERNA
US-09-130-114-1

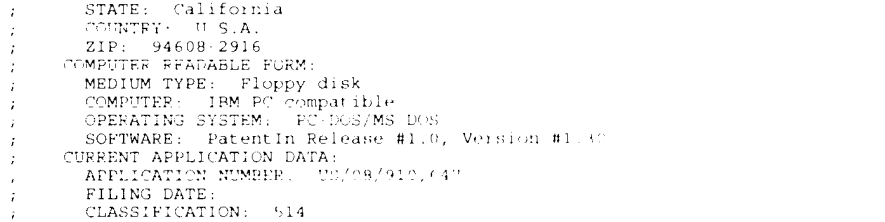
Query Match 1.3%; Score 54.2; DB 2; Length 5452;
Best Local Similarity 45.7%; Pred. No. 0.02;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;





RESULT 11
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match 1.3%; Score 54.2; DB 4; Length 9600;
Best Local Similarity 45.7%; Pred. No. 0.024;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;



```
Db   1253 CAGGAGGGGCAGGAGCAGGAGCAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGG 1312

Qy    328 CAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCA 387

Db   1313 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1372

Qy    388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447

Db   1373 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1432

Qy    448 GCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTG 498

Db   1433 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 1483


RESULT 12
US-09-620-925-1
; Sequence 1, Application US/09620925
; Patent No. 6468986
;  GENERAL INFORMATION:
;       APPLICANT: Zuckermann et al.
;       TITLE OF INVENTION: Compositions and Methods for
;                           Polynucleotide Delivery
;       NUMBER OF SEQUENCES: 4
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Chiron Corporation
;            STREET: 4560 Horton Street
;            CITY: Emeryville
;            STATE: California
;            COUNTRY: U.S.A.
;            ZIP: 94608-2916
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/620,925
;            FILING DATE: 21-Jul-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/910,647
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Fujita, Sharon M.
;            REGISTRATION NUMBER: 38,459
;            REFERENCE/DOCKET NUMBER: 1218.002
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (510) 923 2706
;            TELEFAX: (510) 655-3542
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 9600 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

  Query Match             1.3%;  Score 54.2;  DB 4;  Length 9600;
  Best Local Similarity   45.7%;  Pred. No. 0.024;
  Matches  188;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy      88 CCGGAGAAGCCACAGCGTGACCTGGGATCAGAGGGTGGCTGCCTGAGGGAGACCTCCGAG 147

Db   1073 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGG 1122

Qy     148 GAGGCTGAAGACAGGCCCTATCAGCCCAAGAGCTTCCAGCAGAAACGGAACTATTTCCAG 207

Db   1133 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1182

Qy     208 AGGATGGGGCAGCCACAGATCACAGTGAGGACGATGAAGCCCCTGGCCAAGGTCCACATC 267


Db   1193 CAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGG 1252

Qy     268 CCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGCAGGAG 327

Db   1253 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGCAGGAG 1312

Qy     328 CAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCA 387

Db   1313 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1372

Qy     388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447

Db   1373 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1432

Qy     448 GCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTG 498

Db   1433 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 1483

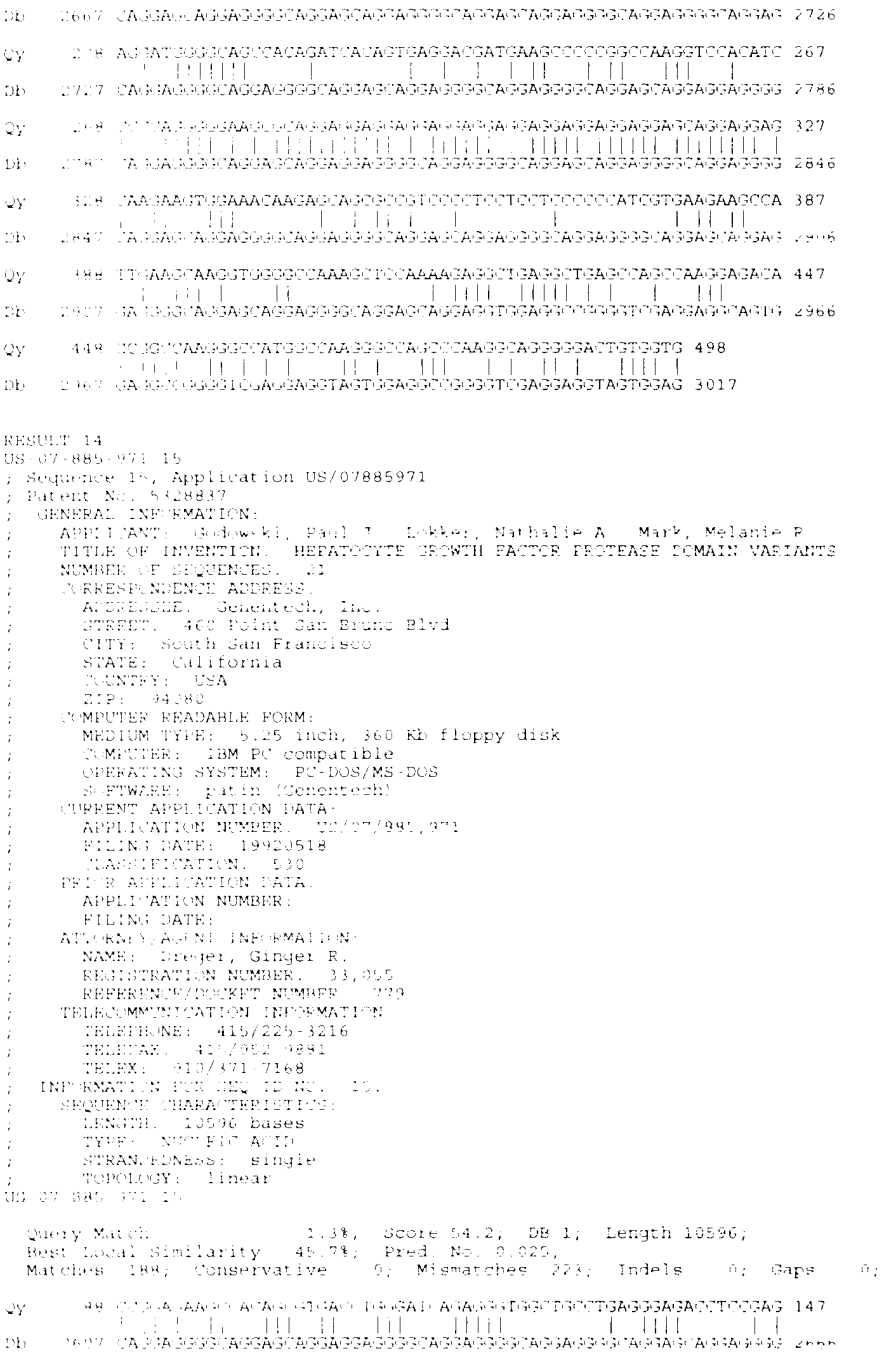

RESULT 13
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/884,811
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-884-811-15

  Query Match             1.3%;  Score 54.2;  DB 1;  Length 10596;
  Best Local Similarity   45.7%;  Pred. No. 0.025;
  Matches  188;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy      88 CCGGAGAAGCCACAGCGTGACCTGGGATCAGAGGGTGGCTGCCTGAGGGAGACCTCCGAG 147

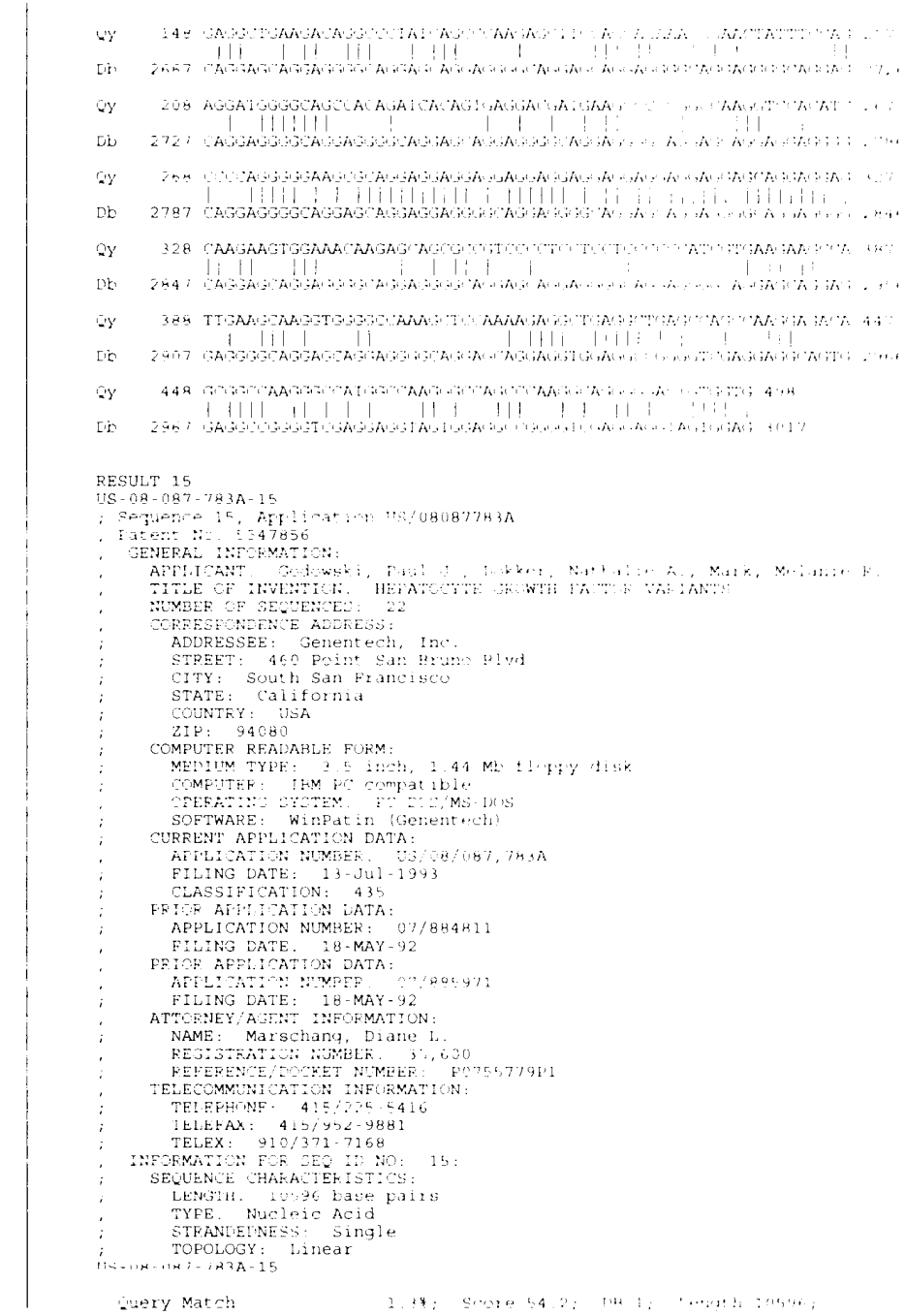
Db   2607 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGG 2656

Qy     148 GAGGCTGAAGACAGGCCCTATCAGCCCAAGAGCTTCCAGCAGAAACGGAACTATTTCCAG 207
```



RESULT 14
US-07-885-971-15
; Sequence 15, Application US/07885971
; Patent No. 5428837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 720
; TELECOMMUNICATION INFORMATION
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13596 bases
; TYPE: Nucleic Acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-885-971-15

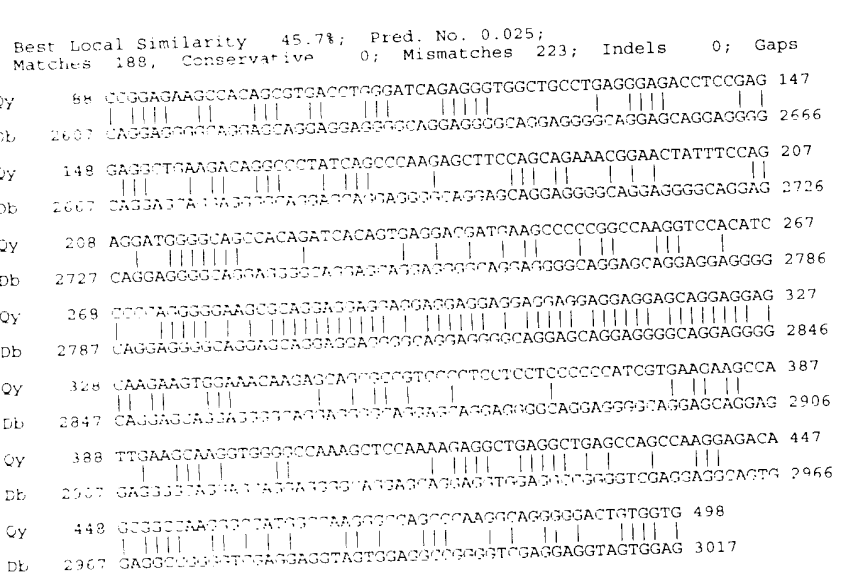
Query Match 1.3%; Score 54.2; DB 1; Length 10596;
Best Local Similarity 45.7%; Pred. No. 0.025;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;



RESULT 15
US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5347856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 2.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,630
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-087-783A-15

Query Match 1.3%; Score 54.2; DB 1; Length 10596;

Best Local Similarity 45.7%; Pred. No. 0.025;
Matches 188, Conservative 0; Mismatches 223; Indels 0; Gaps 0;



Search completed: January 10, 2003, 02:37:11
Job time : 581.572 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic nucleic search, using sw model

Run on: January 9, 2003, 06 18:37 ; Search time 160.711 Seconds
(without alignments)
11430.411 Million cell updates/sec

Title: US-09-803-126-5
Perfect score 4174
Sequence: 1 cggcagcagcaggct.....aaaaaaaaaaaaaaaaaaaa 4174

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4174	100.0	4174	10	US-09-803-126-5	Sequence 5, Appli
2	3352	80.3	3780	10	US-09-803-126-7	Sequence 7, Appli
3	1923.4	46.1	1960	10	US-09-822-849A-265	Sequence 265, App
4	1787.6	42.8	6293	10	US-09-803-126-2	Sequence 2, Appli
5	1232.2	29.5	1247	9	US-09-764-868-145	Sequence 145, App
6	946.4	22.7	4325	10	US-09-803-126-3	Sequence 3, Appli
c 7	715	17.1	24757	9	US-09-764-868-1347	Sequence 1347, Ap
c 8	296	7.1	311	10	US-09-954-456-176	Sequence 176, App
c 9	292	7.0	311	10	US-09-954-456-861	Sequence 861, App
c 10	292	7.0	311	10	US-09-880-107-1520	Sequence 1520, Ap
c 11	196	4.7	526	10	US-09-864-761-7087	Sequence 7087, Ap
c 12	170	4.1	170	10	US-09-864-761-23818	Sequence 23818, A
c 13	149.8	3.6	233	10	US-09-864-761-24717	Sequence 24717, A
c 14	149.8	3.6	566	10	US-09-864-761-8013	Sequence 8013, Ap
15	125.8	3.0	312	10	US-09-782-590-1780	Sequence 1780, Ap
16	102.4	2.5	145	10	US-09-783-590-3077	Sequence 3077, Ap
17	58.2	1.4	3836	12	US-10-002-600-7	Sequence 7, Appli
18	57	1.4	3809	12	US-10-001-870-68	Sequence 68, Appl
c 19	56.4	1.4	659158	9	US-09-771-208-20	Sequence 20, Appl
c 20	55.8	1.3	234	10	US-09-864-761-19061	Sequence 19061, A
c 21	55.8	1.3	468	10	US-09-864-761-2326	Sequence 2326, Ap
c 22	55.4	1.3	267	10	US-09-864-761-27984	Sequence 27984, A
c 23	55.4	1.3	53226	10	US-09-818-264-3	Sequence 3, Appli
24	54.8	1.3	1362	9	US-09-976-740-12	Sequence 12, Appl
25	54.8	1.3	1362	10	US-09-962-055-12	Sequence 12, Appl
26	54.8	1.3	1362	12	US-10-023-529-12	Sequence 12, Appl
27	54.8	1.3	1362	12	US-10-023-523-12	Sequence 12, Appl
28	54.8	1.3	1422	9	US-09-976-740-13	Sequence 13, Appl
29	54.8	1.3	1422	10	US-09-962-055-13	Sequence 13, Appl
30	54.8	1.3	1422	12	US-10-023-529-13	Sequence 13, Appl
31	54.8	1.3	1422	12	US-10-023-523-13	Sequence 13, Appl
32	54.8	1.3	1617	9	US-09-976-740-11	Sequence 11, Appl
33	54.8	1.3	1617	10	US-09-962-055-11	Sequence 11, Appl
34	54.8	1.3	1617	12	US-10-023-529-11	Sequence 11, Appl
35	54.8	1.3	1617	12	US-10-023-523-11	Sequence 11, Appl
36	54.8	1.3	2561	9	US-09-976-740-48	Sequence 48, Appl
37	54.8	1.3	2561	10	US-09-962-055-48	Sequence 48, Appl
38	54.8	1.3	2561	12	US-10-023-523-48	Sequence 48, Appl
c 39	54.6	1.3	474	10	US-09-864-761-11284	Sequence 11284, A
c 40	54	1.3	465237	10	US-09-933-267A-1	Sequence 1, Appli
c 41	53.4	1.3	360	10	US-09-864-761-28274	Sequence 28274, A
c 42	53.4	1.3	450	10	US-09-864-761-11696	Sequence 11696, A
c 43	53.4	1.3	475	10	US-09-864-761-6203	Sequence 6203, Ap
c 44	53.4	1.3	511	10	US-09-864-761-22817	Sequence 22817, A
45	53.4	1.3	4744	10	US-09-880-107-3459	Sequence 3459, Ap

ALIGNMENTS

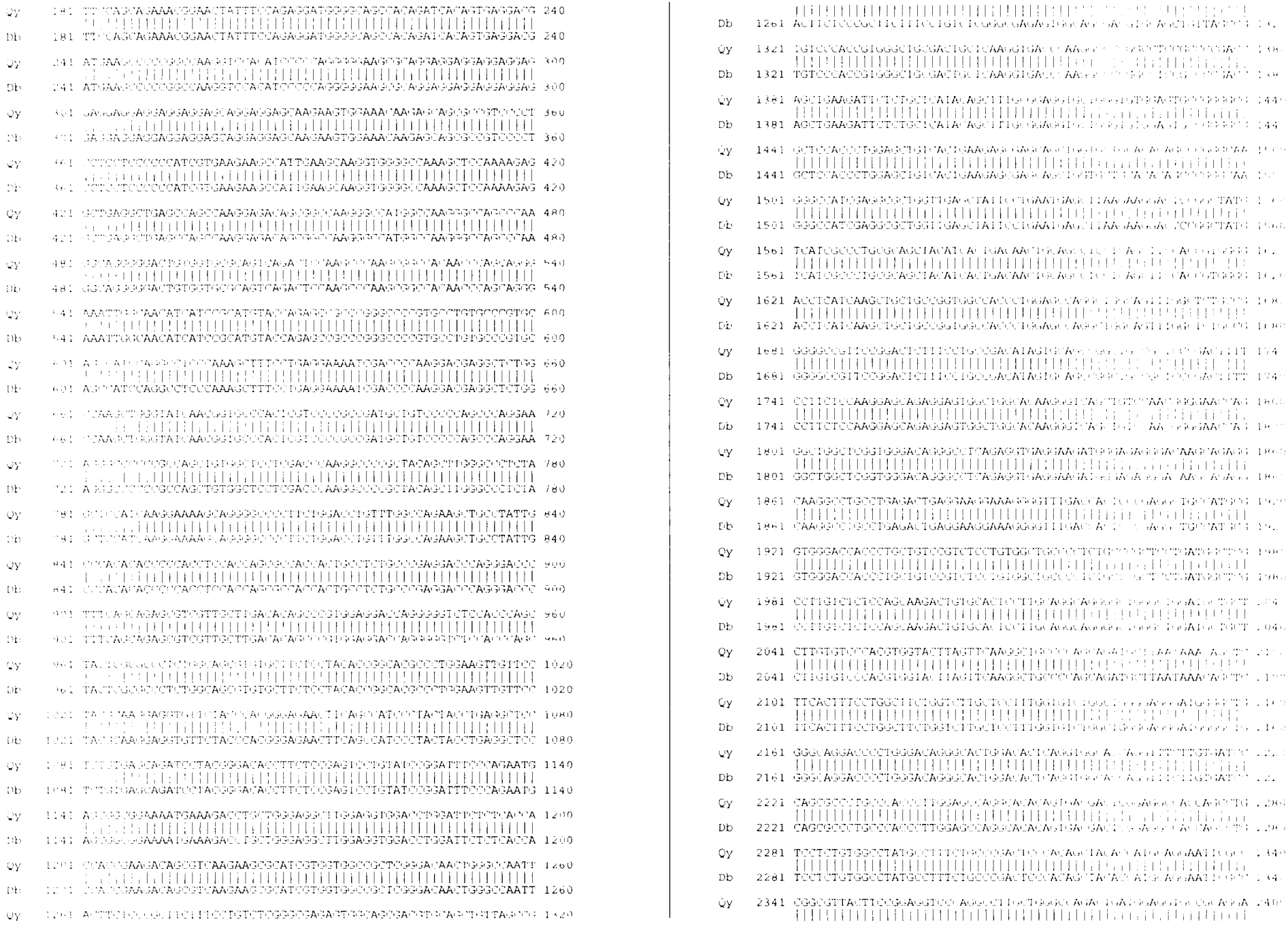
RESULT 1
US-09-803-126-5
; Sequence 5, Application US/09803126
; Patent No US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional My sin Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related
; OTHER INFORMATION: protein variant 1 (hMRP1)
US-09-803-126-5

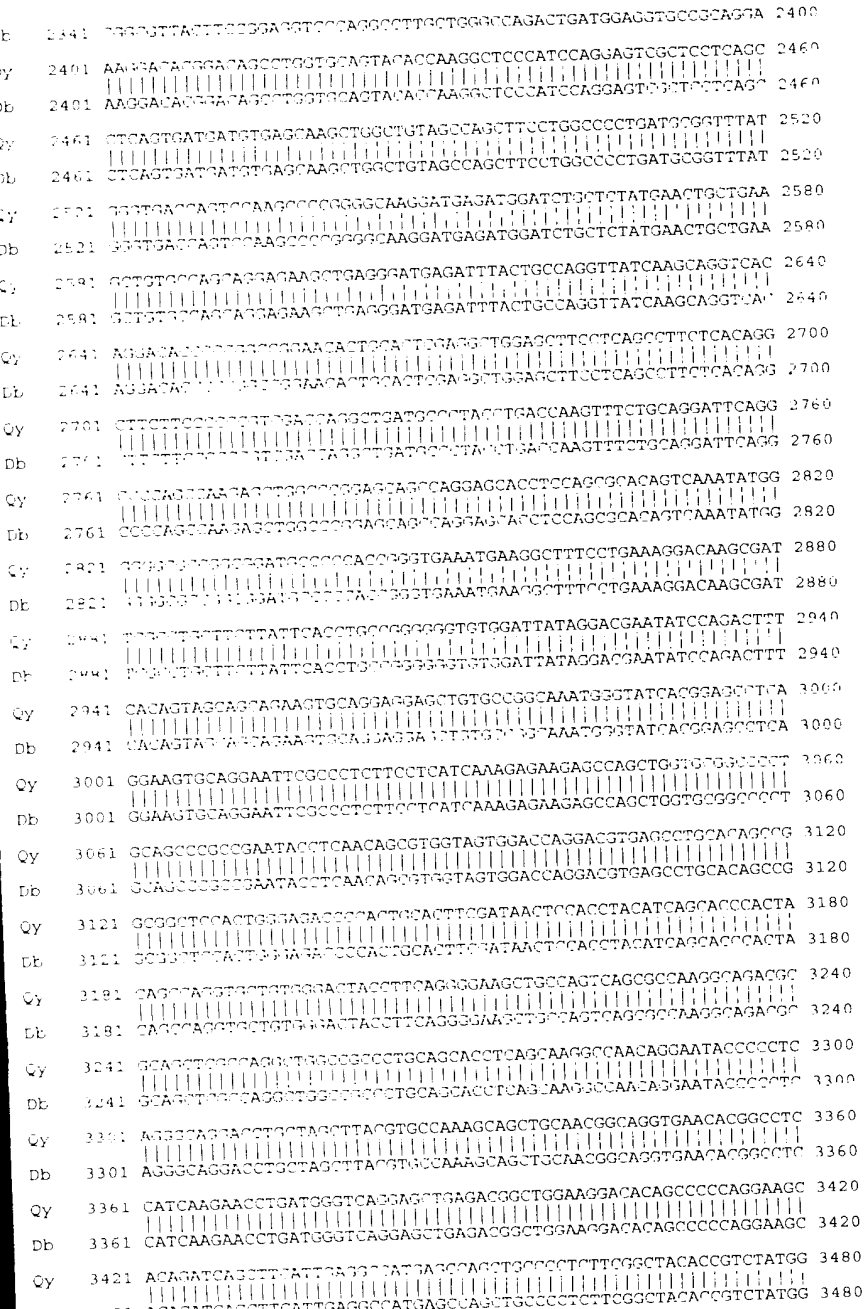
Query Match 100.0%; Score 4174; DB 10; Length 4174;
Best Local Similarity 100.0%; Pred. No. 0,
Matches 4174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

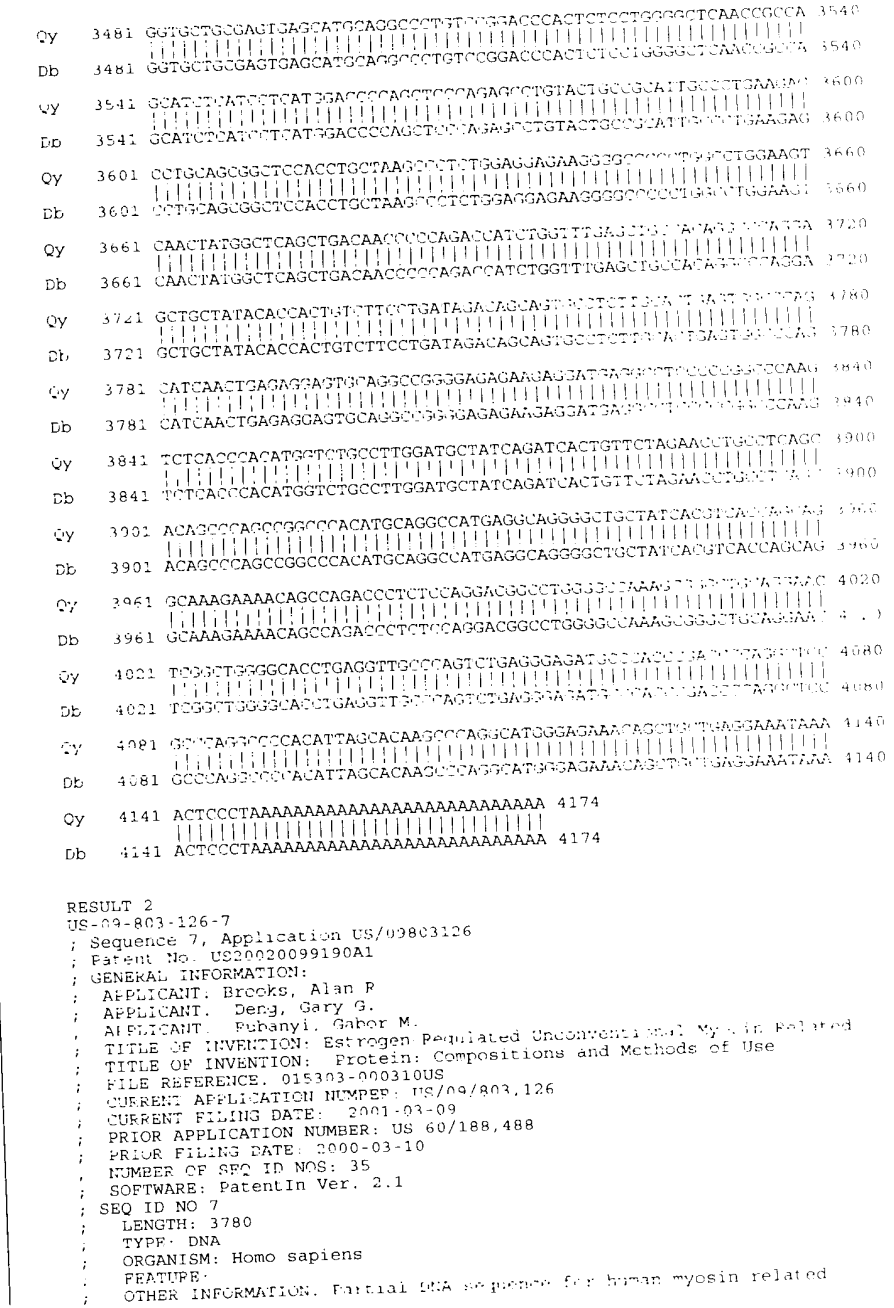
Qy 1 CGGCAGCAGCAGGCTCGGGCCTCCGAGGCTGCGTCCCAGGCCTCAGCCTCAGCCGTCACC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CGGCAGCAGCAGGCTCGGGCCTCCGAGGCTGCGTCCCAGGCCTCAGCCTCAGCCGTCACC 60

Qy 61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCCTGACCTGCGATCAGAG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCCTGACCTGCGATCAGAG 120

Qy 121 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180



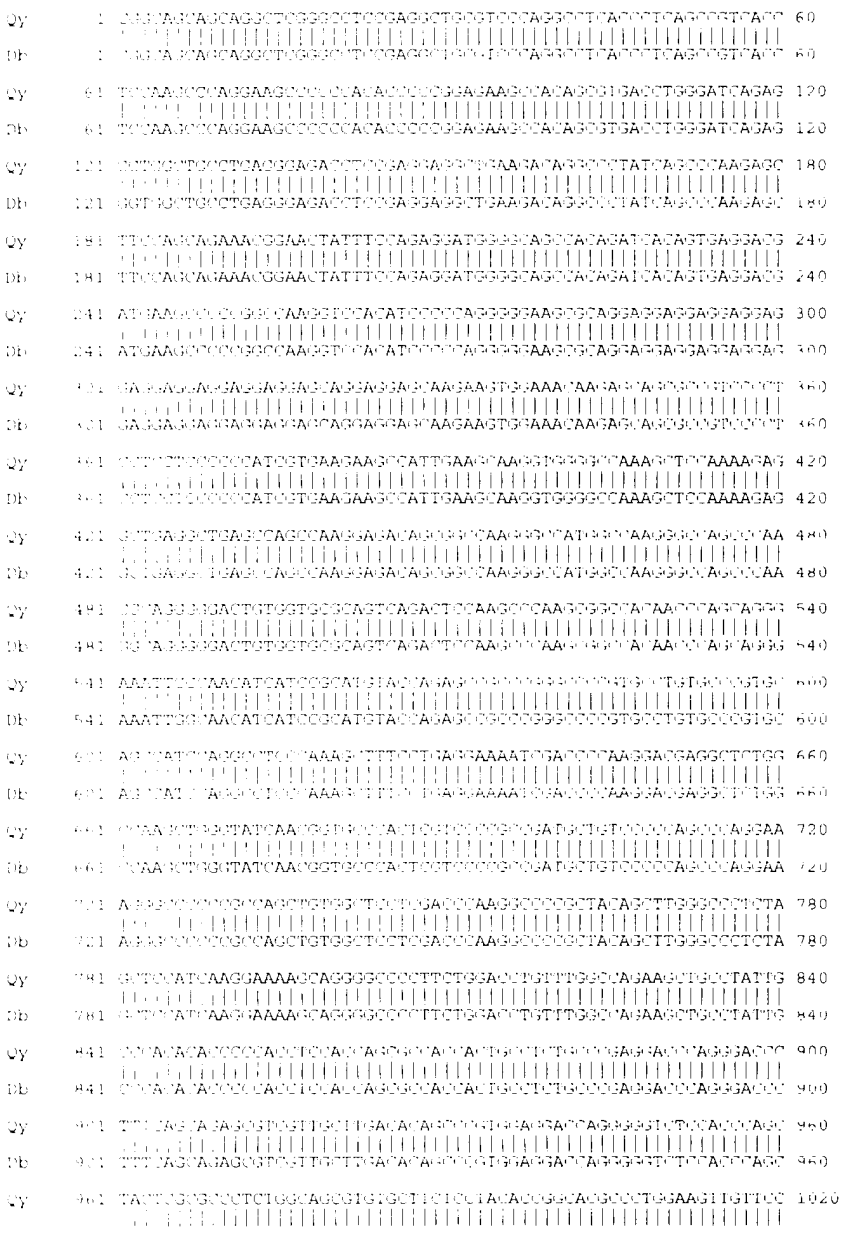


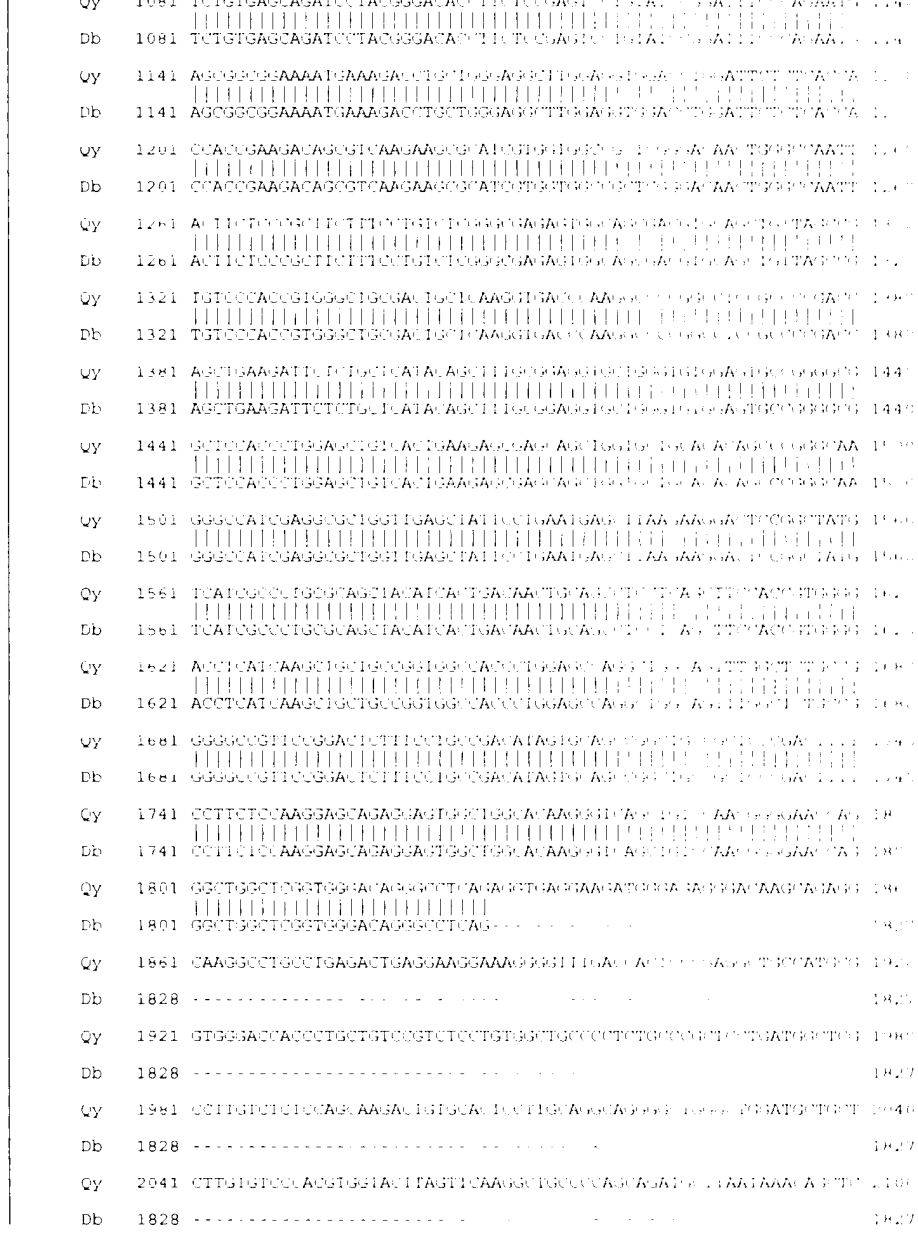


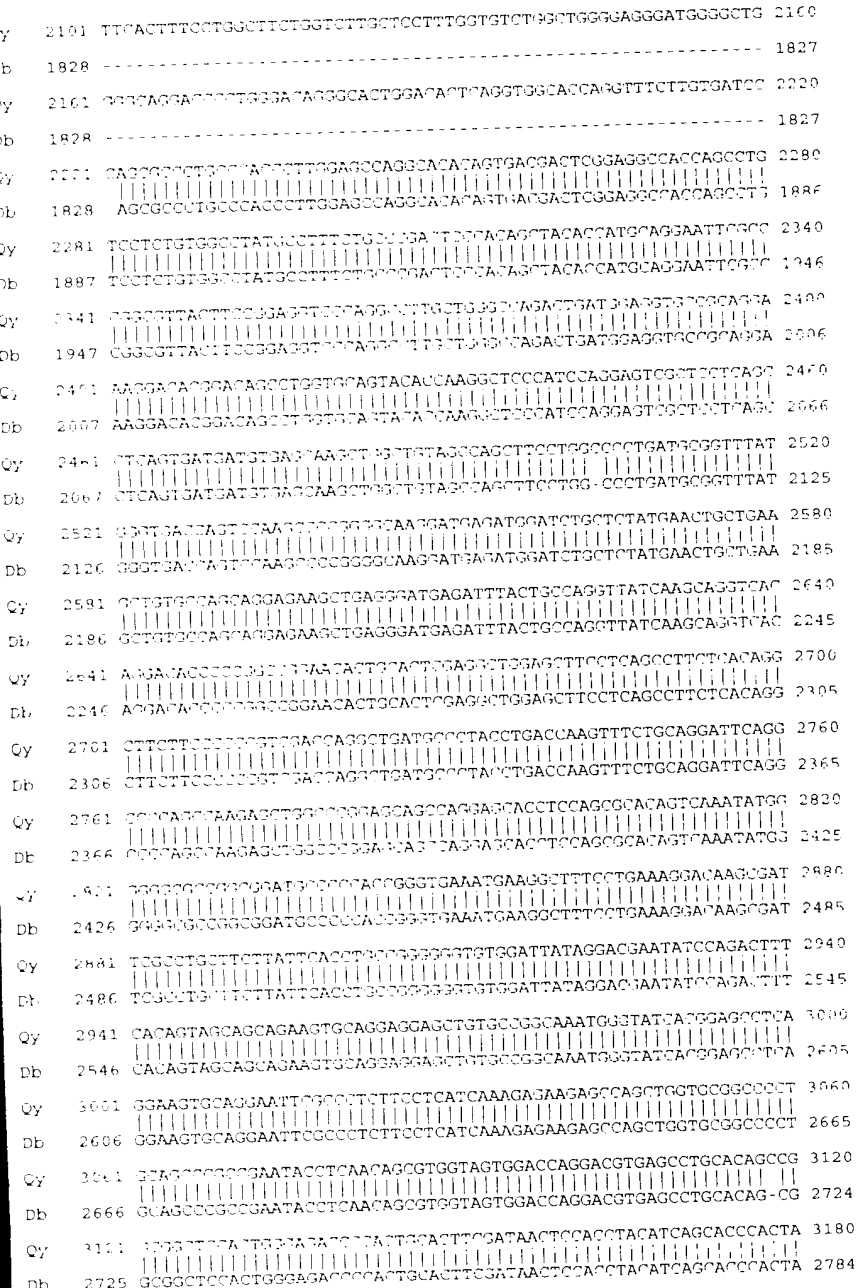
RESULT 2
US-09-803-126-7
; Sequence 7, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R
; APPLICANT: Deng, Gary G.
; APPLICANT: Fuhanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related

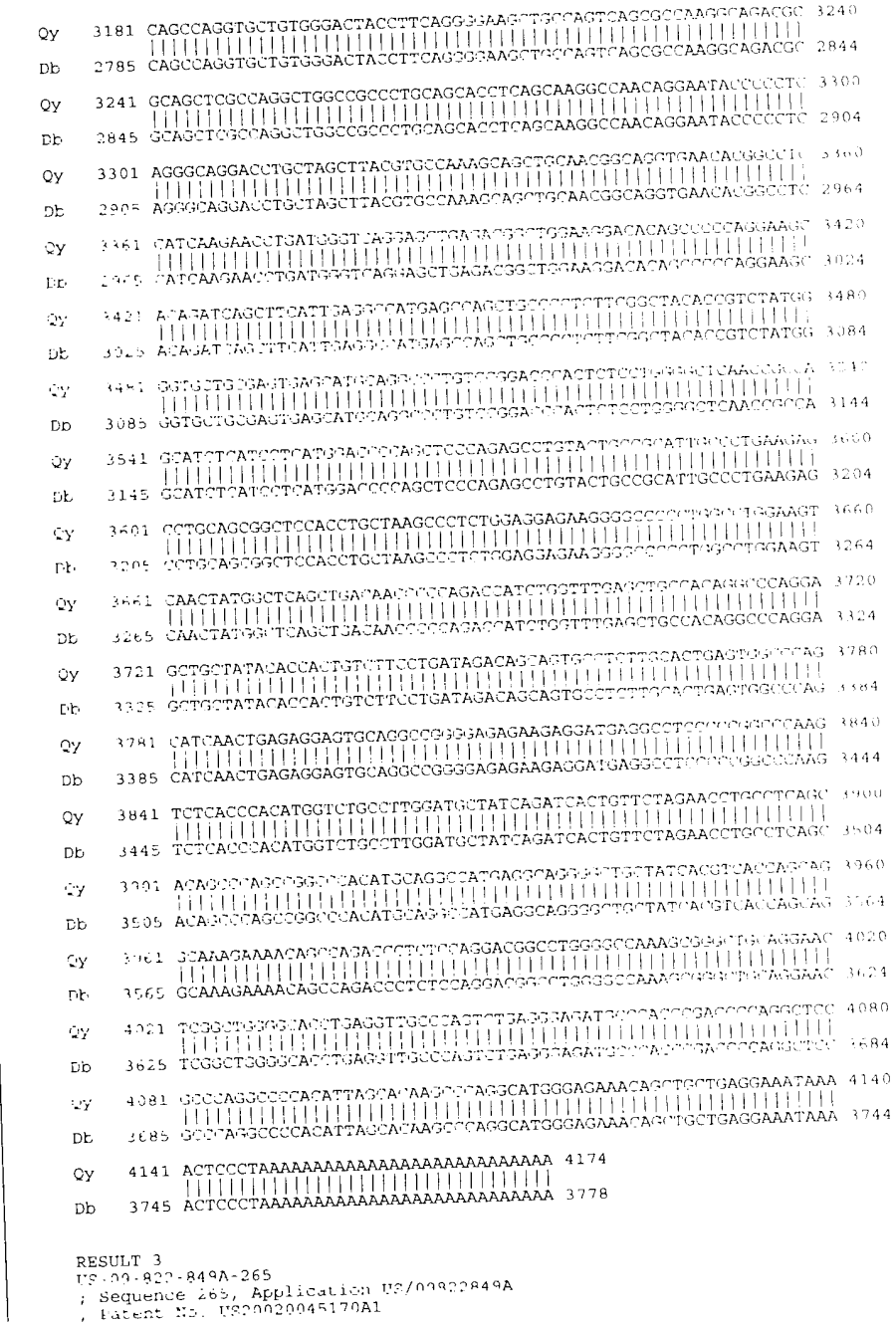
; OTHER INFORMATION: protein variant 2 (hMMP2)
OS US-09-803-126-5

Query Match 80.3%; Score 3352; DB 10; Length 3780;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 3379; Conservative 0; Mismatches 0; Indels 396; Gaps 3;









RESULT 3
US-09-820-849A-265
; Sequence 265, Application US/09820849A
; Patent No. US20020045170A1

```
; GENERAL INFORMATION:
;  APPLICANT: Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Pechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 0403
;  CURRENT APPLICATION NUMBER: US/09/822,849A
;  CURRENT FILING DATE:  2001-09-04
;  PRIOR APPLICATION NUMBER: 60/195,582
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 598
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 265
;  LENGTH: 1960
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-822-849A-265

  Query Match             46.2%;  Score 1923.4;  DB 10;  Length 1960;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1910;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy    2192 CTTAATTGGCACTAGGTTTCTTGTGGATCCCAGCGGCCTGCCCACCCTTGGAGCCAGGCAC 2251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCAGTTGGCACCAGGTTTCTTGTGGATCCCAGCGGCCTGCCCACCCTTGGAGCCAGGCAC 60

Qy    2252 ATAGTGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACT 2311
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACAGTGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACT 120

Qy    2312 CTTATAGCTACCACATGCAGGAATTCGCCCGGCGTTACTTCCGGAGGTCCCAGGCCTTGC 2371
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTTATAGCTACCACATGCAGGAATTCGCCCGGCGTTACTTCCGGAGGTCCCAGGCCTTGC 180

Qy    2372 TGGGCCAGACTGATGGAGGTGCCGCAGGAAAGGACACGGACAGCCTGGTGCAGTACACCA 2431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGGGCCAGACTGATGGAGGTGCCGCAGGAAAGGACACGGACAGCCTGGTGCAGTACACCA 240

Qy    2432 ACCCTCCCATCCAGGAGTCGCTCCTCAGCCTCAGTGATGATGTGAGCAAGCTGGCTGTAG 2491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACCCTCCCATCCAGGAGTCGCTCCTCAGCCTCAGTGATGATGTGAGCAAGCTGGCTGTAG 300

Qy    2492 CCATCTTTCCTGGCCCCTGATGCGGTTTATGGGCAGCTACTCCAAGCCCCGGGGCAAGGAT 2551
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCATCTTTCCTGGCCCCTGATGCGGTTTATGGGCAGCTACTCCAAGCCCCGGGGCAAGGAT 359

Qy    2552 GAGATGGATCTGCTCTATGAACTGCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGATGAG 2611
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 GAGATGGATCTGCTCTATGAACTGCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGATGAG 419

Qy    2612 ATTAATGCAGGTTATCAAGAGGTCACAGAACTACCCCCGGCCGGAACACTGCACTCGA 2671
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 ATTAATGCAGGTTATCAAGAGGTCACAGGACTACCCCCGGCCGGAACACTGCACTCGA 479

Qy    2672 GCCCAAGCTTCCTCAGCCTTCTCACAGGCTTCTTCCCCCGTGGACCAGGCTGATGCCC 2731
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 GCCCAAGCTTCCTCAGCCTTCTCACAGGCTTCTTCCCCCGTGGACCAGGCTGATGCCC 539

Qy    2732 TACCTGACCAAGTTTCTGCAGGATTCAGGCCCAGCCAAGAGCTGGCCCGGAGCAGCCAG 2791
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 TACCTGACCAAGTTTCTGCAGGATTCAGGCCCAGCCAAGAGCTGGCCCGGAGCAGCCAG 599

Qy    2792 GAGCACTTCCTAGCGCACAGTCAAATATGGGGGGCGCCGGCGGATGCCCCCACGGGTGAA 2851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GAGCACTTCCTAGCGCACAGTCAAATATGGGGGGCGCCGGCGGATGCCCCCACGGGTGAA 659

Qy    2852 ATTAAAGCCTTTCCTGAAAGGAGCAAGCGAATCGGCCTGCTTCTTATTCACCTGCCGGGGGGT 2911
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATGAAGGCTTTCCTGAAAGGAGCAAGCGAATCGGCCTGCTTCTTATTCACCTGCCGGGGGGT 719

Qy    2912 GTGGATTATAGGACGAATATCCAGACTTTCACAGTACGCAGTCAGAATCTCAGCAGAGTCT 2971
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 GTGGATTATAGGACGAATATCCAGACTTTCACAGTACGCAGTCAGAATCTCAGCAGAGTCT 779

Qy    2972 TGCCCGCAAATGGGTATCACGGAGCCTCAGGAAGTGCAGAATTCCCTCTTCTCTATTCAT 3031
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TGCCCGCAAATGGGTATCACGGAGCCTCAGGAAGTGCAGAATTCCCTCTTCTCTATTCAT 839

Qy    3032 AAAGAGAAGAGCCAGCTGGTGCCGGCCCTGCAGCCCCCGGAATACCTCAAACAGCCTCTA 3091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AAAGAGAAGAGCCAGCTGGTGCCGGCCCTGCAGCCCCCGGAATACCTCAAACAGCCTCTA 899

Qy    3092 GTGGACCAGGACCTGAGCCTGCACAGCCCGGCGCCCACCCCAGAGCCCACTCCTATTCT 3151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GTGGACCAGGACCTGAGCCTGCACAGCCCGGCGCCCACCCCAGAGCCCACTCCTATTCT 959

Qy    3152 GATAACTCCCACCTACATCAGCACCCACTACAGCACAGGTCCTCCAGACTACTTCAAGGAG 3211
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 GATAACTCCCACCTACATCAGCACCCACTACAGCACAGGTCCTCCAGACTACTTCAAGGAG 1019

Qy    3212 AAGCTGCCAGTCAGCGCCAAGGCAGAGGCGTAGCTCCGCAAGCCGCCTCTGCAGCAC 3271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 AAGCTGCCAGTCAGCGCCAAGGCAGAGGCGTAGCTCCGCAAGCCGCCTCTGCAGCAC 1079

Qy    3272 CTCAGCAAGGCCAACAGGAATACCCCTCAGGGCAGGACCGCCAGCTACTGCCGCAAAG 3331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 CTCAGCAAGGCCAACAGGAATACCCCTCAGGGCAGGACCGCCAGCTACTGCCGCAAAG 1139

Qy    3332 CAGCTGCAACGGCAGGTGAACACGGCCTCCATCAAGAACCTGATGGCCAGGAGCTGGAGA 3391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 CAGCTGCAACGGCAGGTGAACACGGCCTCCATCAAGAACCTGATGGCCAGGAGCTGGAGA 1199

Qy    3392 CGGCTGGAAGGACACAGCCCCCAGGAAGCACAGATCAGCTTCATGAAGCCATGAGCCAG 3451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 CGGCTGGAAGGACACAGCCCCCAGGAAGCACAGATCAGCTTCATGAAGCCATGAGCCAG 1259

Qy    3452 CTGCCCCTCTTCGGCTACACCGTCTATGGGGTGCTGCGAGTGAGCATCAAGGCCTATGTC 3511
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 CTGCCCCTCTTCGGCTACACCGTCTATGGGGTGCTGCGAGTGAGCATCAAGGCCTATGTC 1319

Qy    3512 GGACCCACTCTCCTGGGGCTCAACGGCGTAGCATCTCATCTCATTGAACCCAGCTCTGTAG 3571
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 GGACCCACTCTCCTGGGGCTCAACGGCGTAGCATCTCATCTCATTGAACCCAGCTCTGTAG 1379

Qy    3572 AGCCTGTACTGCCGCATTGCCCTGAAGAGCCTGCAGGGGTCCCACTTGTAAGCCCCTG 3631
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 AGCCTGTACTGCCGCATTGCCCTGAAGAGCCTGCAGGGGTCCCACTTGTAAGCCCCTG 1439

Qy    3632 GAGGAGAAGGGGCCCCCTGGCCTGGAAGTCAACTATGGCTCAGCCGACAACCCCAGACC 3691
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 GAGGAGAAGGGGCCCCCTGGCCTGGAAGTCAACTATGGCTCAGCCGACAACCCCAGACC 1499

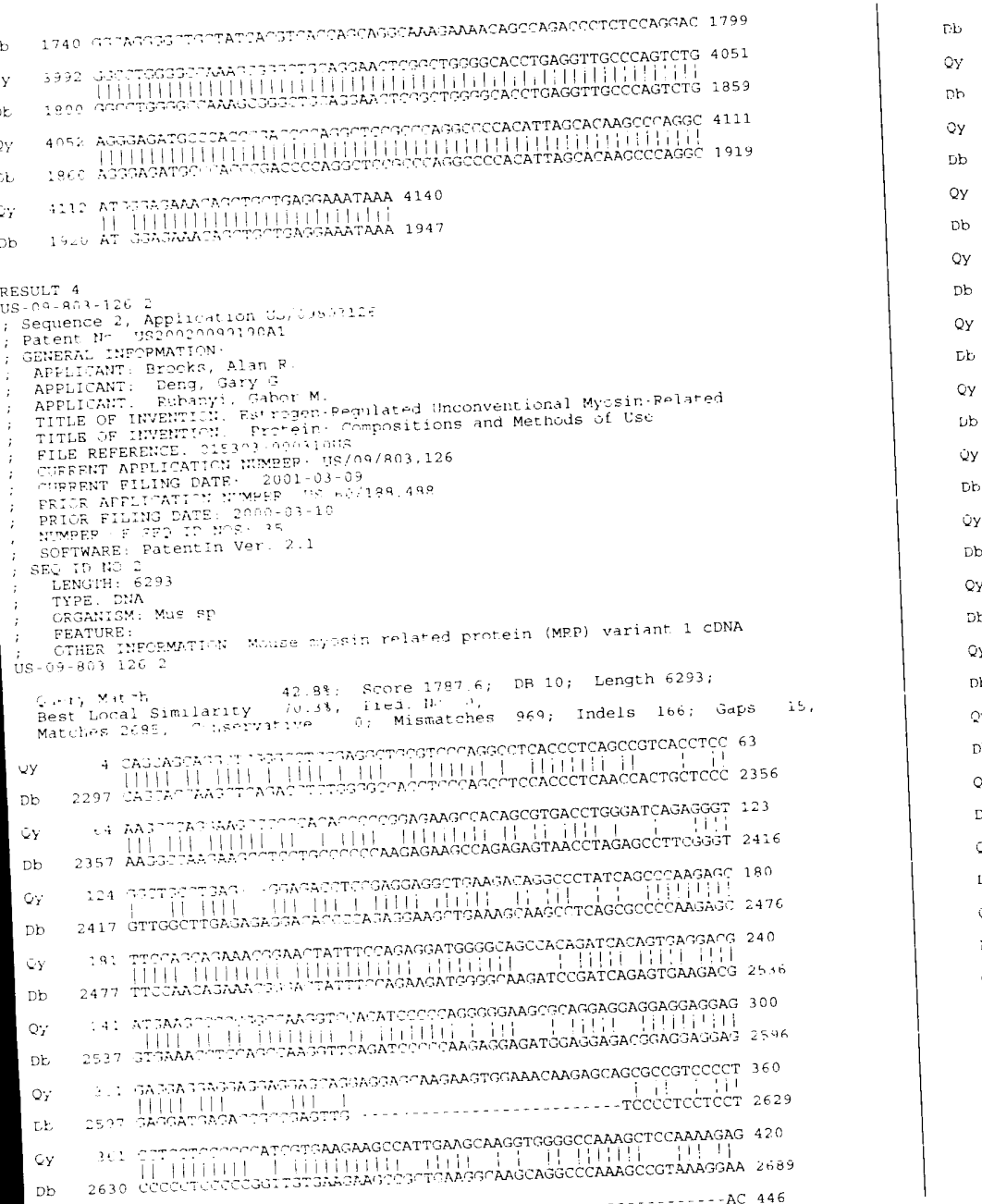
Qy    3692 ATCTGGTTTGAGCTGCCACAGGCCCAGGAGCTGCTATACACCACTTTTCTTCATACAC 3751
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 ATCTGGTTTGAGCTGCCACAGGCCCAGGAGCTGCTATACACCACTTTTCTTCATACAC 1559

Qy    3752 AGCAGTGCCTCTTGCACTGAGTGGCCCAGCATCAACTGAGAGGAGCCGCAGGCGGGAGA 3811
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 AGCAGTGCCTCTTGCACTGAGTGGCCCAGCATCAACTGAGAGGAGCCGCAGGCGGGAGA 1619

Qy    3812 GAAGAGGATGAGGCCTCCCCCGGCCCAAGTCTCACCCACATGGCCCTTTTGATGCTAT 3871
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GAAGAGGATGAGGCCTCCCCCGGCCCAAGTCTCACCCACATGGCCCTTTTGATGCTAT 1679

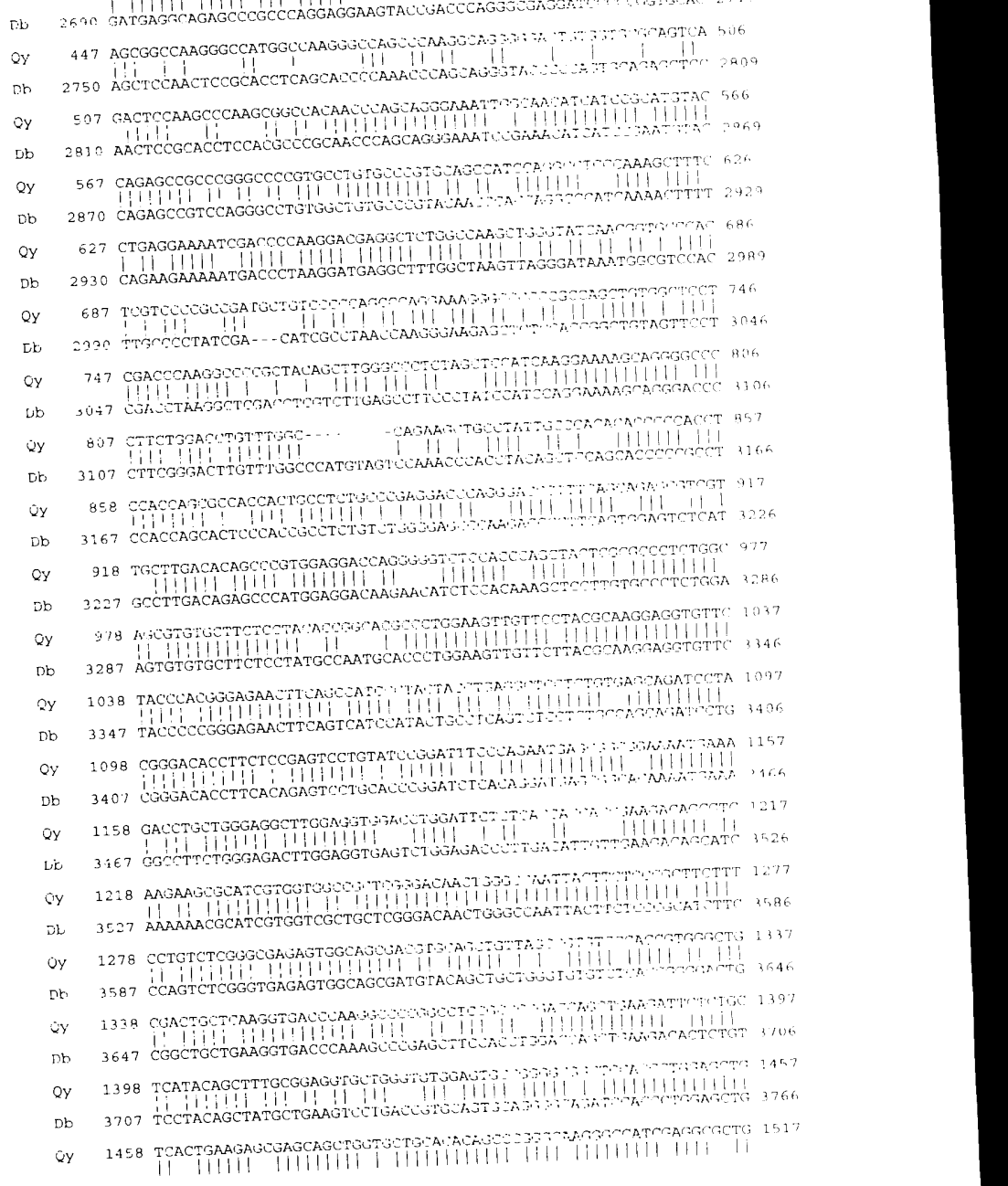
Qy    3872 CAGATCACTGTTCTAGAACCTGCCTCAGCACAGCCCAGCCGGCCTACATGTAGGCCATGA 3931
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 CAGATCACTGTTCTAGAACCTGCCTCAGCACAGCCCAGCCGGCCTACATGTAGGCCATGA 1739

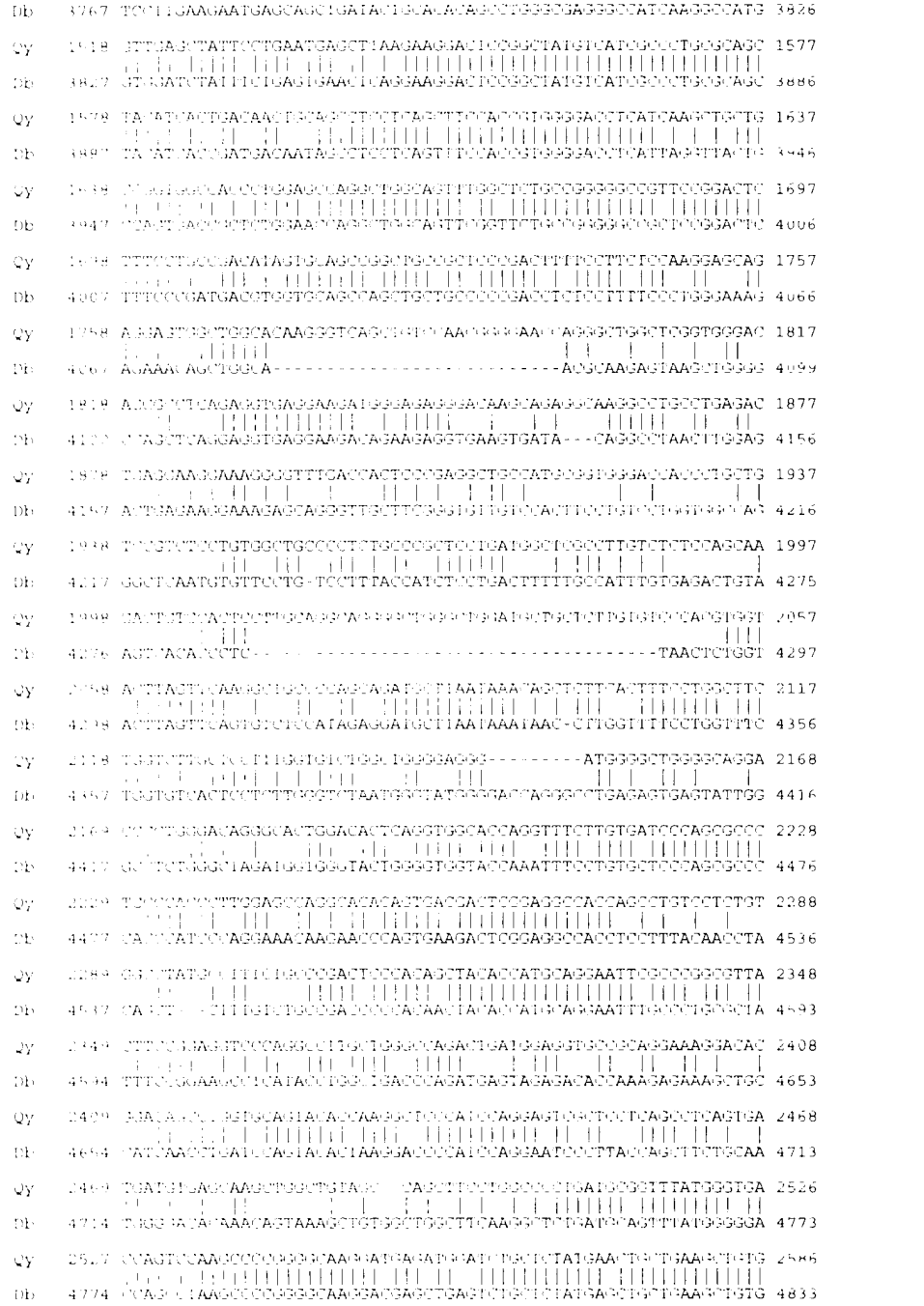
Qy    3932 GGCAGGGGCTGCTATCAGGTCACCAGCAGGCAAAGAAAAAGCCAAAGAAAGTCCAGA 3991
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

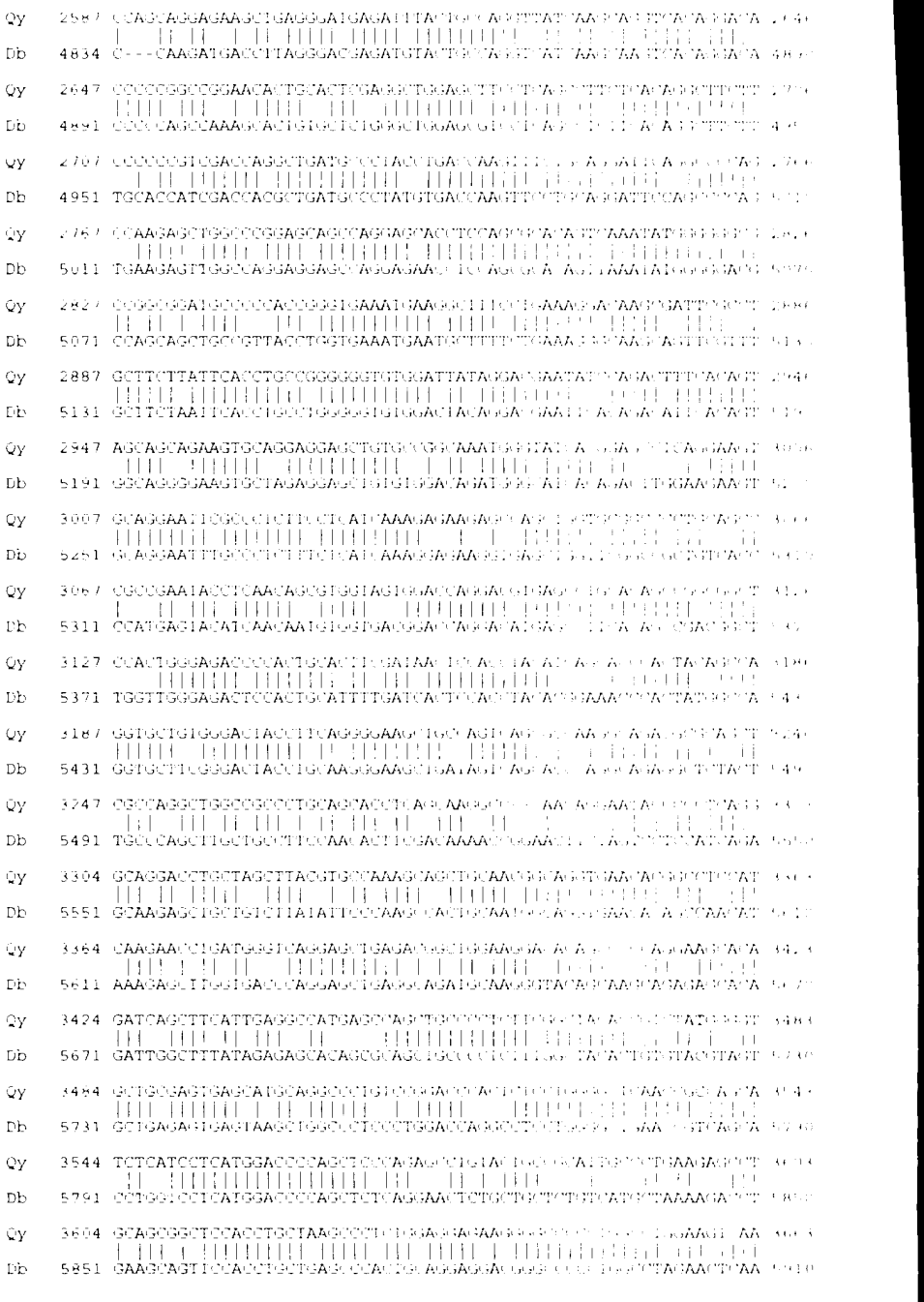


RESULT 4
US-09-803-126-2
; Sequence 2, Application US/09803126
; Patent No. US20020090610A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein Compositions and Methods of Use
; FILE REFERENCE: 015303-000410US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,498
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 6293
; TYPE: DNA
; ORGANISM: Mus sp
; FEATURE:
; OTHER INFORMATION: Mouse myosin related protein (MRP) variant 1 cDNA
US-09-803-126-2

Query Match 42.8%; Score 1787.6; DB 10; Length 6293;
Best Local Similarity 70.3%; Pred. No. 0.;
Matches 2685; Conservative 0; Mismatches 969; Indels 166; Gaps 15;







```
Qy    3664 CTATGGCTCAGTTGACAAGCCCCAGACCATCTGGTTTGAGCTGCCACAGGCCCAGGAGCT 3723
           ||||||||| | |||||||||||||||||||||||||||||||||||||||||||||||
Db    5911 CTATGGCTCTGTTGACAAGCCCCAGACCATCTGGTTTGAGTTGCCACAGGCCCAGGAGCT 5970

Qy    3724 GCTATACACCACTGTCTTCCTGATAGACAGCAGTGGCTCT 3763
           ||  ||||||   |||||||| | ||||||   || ||
Db    5971 GCAGCACACAATATCTTCCTGCTGGGCAGCATGTCCACT 6010


RESULT 5
US-09-764-868-145
; Sequence 145, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PT032
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 145
;  LENGTH: 1247
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-764-868-145

  Query Match             29.5%;  Score 1232.2;  DB 9;  Length 1247;
  Best Local Similarity   99.4%;  Pred. No. 7.3e-247;
  Matches 1231;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     872 ACTGGCTTTGCCCGAGGACCCAGGAACCTTTCAGCAGAGCGTCGTTGCTTGACACAGCC 931
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACTGGCTTTGCCCGAGGACCCAGGGACCTTTCAGCAGAGCGTCGTTGCTTGACACAGCC 60

Qy     932 CGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGCGCCCTCTGGCAGCGTGTGCTTCTC 991
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGCGCCCTCTGGCAGCGTGTGCTTCTC 120

Qy     992 CTACACCGGCAGGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGGGAGAA 1051
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTACACCGGCAGGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTCTACCCACGGGAGAA 180

Qy    1052 CTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTC 1111
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTC 240

Qy    1112 CGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGG 1171
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGG 300

Qy    1172 CTTGGAGGTGGACCTGGATTCTCTCACCACCACCGAAGACAGCGTCAAGAAGCGCATCGT 1231
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTTGGAGGTGGACCTGGATTCTCTCACCACCACCGAAGACAGCGTCAAGAAGCGCATCGT 360

Qy    1232 GGTGGCGGCTCGGGACAACTGGGCCAATTACTTCTCCCGCTTCTTTCCTGTCTCGGGCGA 1291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTGGCGGCTCGGGACAACTGGGCCAATTACTTCTCCCGCTTCTTTCCTGTCTCGGGCGA 420

Qy    1292 GAGTGGCAGCGACGTCCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTCAAGGT 1351
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAGTGGCAGCGACGTCCAGCTGTTAGCCGTGTCCCACCGTGGGCTGCGACTGCTCAAGGT 480

Qy    1352 GACCCAAGGCCCCGTCCTCCGCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 1411
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GACCCAAGGCCCCGTCCTCCGCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 540

Qy    1412 GGAGGTGCTGGGTGTGGAGTGCGGGGGGGGCTCCACCCTGGAGCTGTCACTGAAGAGCGA 1471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGAGGTGCTGGGTGTGGAATGCGGGGGGGGCTCCACCCTGGAGCTGTCACTGAAGAGCGA 600

Qy    1472 GCAGCTGGTGCTGCACACAGCCCGGGCAAGGGCCATCGAGGCGCTGGTTGAGCTATTCCT 1531
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GCAGCTGGTGCTGCACACAGCCCGGGCAAGGGCCATCGAGGCGCTGGTTGAGCTATTCCT 660

Qy    1532 GAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAA 1591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAACGAGCTTAAGAAGGACTCCGGCTATGTCATCGCCCTGCGCAGCTACATCACTGACAA 720

Qy    1592 CTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTTGTTGCCGTGGCCACCCT 1651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTTGTTGCCGTGGCCACCCT 780

Qy    1652 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACAT 1711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTCTTTCCTGCCGACAT 840

Qy    1712 AGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 1771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCA 900

Qy    1772 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGT 1831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGT 960

Qy    1832 GAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGG 1891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GAGGAAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGG 1020

Qy    1892 GGTTTGACCACTCCCGAGGCTGCCATGCGGTGGGACCACCCTGCTGTCCGTCTCCTGTGG 1951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGTTTGACCACTCCCGAGGCTGCCATGCGGTGGGACCACCCTGCTGTCCGTCTCCTGTGG 1080

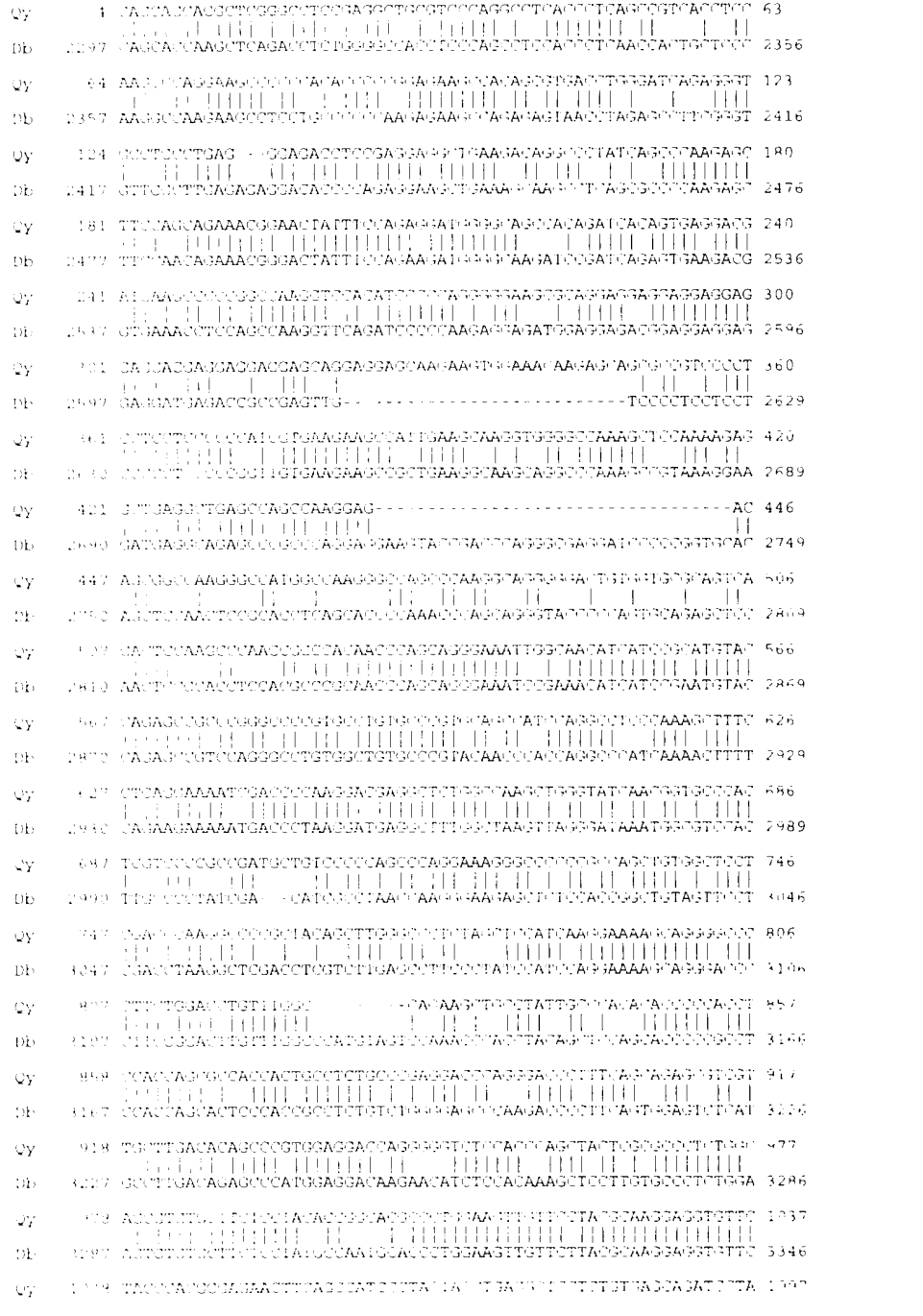
Qy    1952 CTGCCCCTCTGCCCGCTCCTGATGGCTTGCCTTGTCTCTCCAGCAAGACTGTGCACTCCT 2011
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CTGCCCCTCTGCCCGCTCCTGATGGCTTGCCTTGTCTCTCCAGCAAGACTGTGCACTCCT 1140

Qy    2012 TGCAGGCAGGGGCTGGGCTGGPTGCTGCTTTGTGTCCCACGGTACTTAGTTCAAGGC 2071
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGCAGGCAGGGGCTGGGCTGGPTGCTGCTTTGTGTCCCACGGTACTTAGTTCAAGGC 1200

Qy    2072 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 2108
           |||||||||||||||||||||||||||||||||||||
Db    1201 TGCCCCAGCAGATGCTTAATAAACAGCTCTTCACTTT 1237


RESULT 6
US-09-803-126-3
; Sequence 3, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
;  APPLICANT: Brooks, Alan R.
;  APPLICANT:  Deng, Gary G.
;  APPLICANT:  Fabanyi, Gabor M.
;  TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
;  TITLE OF INVENTION:  Protein: Compositions and Methods of Use
;  FILE REFERENCE: 015303-000310US
;  CURRENT APPLICATION NUMBER: US/09/803,126
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: US 60/188,488
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;  LENGTH: 4375
;  TYPE: DNA
;  ORGANISM: Mus sp.
;  FEATURE
;  OTHER INFORMATION: Mouse myosin related protein (MRP) variant 2 cDNA
US-09-803-126-3

  Query Match             22.7%;  Score 946.4;  DB 10;  Length 4375;
  Best Local Similarity   72.9%;  Pred. No. 2.4e-187;
  Matches 1322;  Conservative    0;  Mismatches  416;  Indels   76;  Gaps    5;
```



RESULT 7
US-09-764-868-1347/c
; Sequence 1347, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1347
; LENGTH: 24757
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1347

Query Match 17.1%; Score 715; DB 9; Length 24757;
Best Local Similarity 99.8%; Pred. No. 4.5e-139;

Matches 718; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy   1656 CCAGGCTTGGCAGTTTGGCTCTGCCGGGGGCGGTTCCGGACTCTTTCCTGCCGACATAGTG 1715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10170 CCAGGCTTGGCAGTTTGGCTCTGCCGGGGGCGGTTCCGGACTCTTTCCTGCCGACATAGTG 10111

Qy   1716 CAGCCGGCTGCCGTTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAG 1775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10110 CAGCCGGCTGCCGTTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAG 10051

Qy   1776 GGTCAGGTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGG 1835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10050 GGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGG 9991

Qy   1836 AAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGGGGTT 1895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9990 AAGATGGGAGAGGGACAAGCAGAGGCAAGGCCTGCCTGAGACTGAGGAAGGAAAGGGGTT 9931

Qy   1896 TGACCACTCCCCAGGCCTCCCATCCGGTGGGACCACCCTGCTGTCCGTCTCCTGTGGCTGC 1955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9930 TGACCACTCCCCAGGCCTCCCATCCGGTGGGACCACCCTGCTGTCCGTCTCCTGTGGCTGC 9871

Qy   1956 CCCTCTTCCCCCCTCTTGATGCCTCCCCTTGTCTCTCCAGCAAGACTGTGCACTCCTTGCA 2015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9870 CCCTCTTCCCCCCTCTTGATGGCTCCCCTTGTCTCTCCAGCAAGACTGTGCACTCCTTGCA 9811

Qy   2016 GGCAGGGCTCCCTTGGATGCTGCTCTTGTGTCCCACGTGGTACTTAGTTCAAGGCTGCC 2075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9810 GGCAGGGCTCCCTTGGATGCTGCTCTTGTGTCCCACGTGGTACTTAGTTCAAGGCTGCC 9751

Qy   2076 CCAGTAGATGTTTAATAAACAGCTCTTCACTTTCCTGGCTTCTGGTCTTGCTCCTTTGGT 2135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9750 CCAGCAGATGTTTAATAAACAGCTCTTCACTTTCCTGGCTTCTGGTCTTGCTCCTTTGGT 9691

Qy   2136 GCTTGGTTGGGAAGTATGGGCTGGGGCAGGACCCCTGGGACAGGGCACTGGACACTCA 2195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9690 GCTTGGTTGGGAAGTATGGGCTGGGGCAGGACCCCTGGGACAGGGCACTGGACACTCA 9631

Qy   2196 GGTGGCACCAGGTTTCTTGTGATCCCAGGGCCCTGCCCACCCTTGGAGCCAGGCACACAG 2255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9630 GGTGGCACCAGGTTTCTTGTGATCCCAGGGCCCTGCCCACCCTTGGAGCCAGGCACACAG 9571

Qy   2256 TGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCA 2315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9570 TGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCA 9511

Qy   2316 CAGCTACACCATGCAGGAATTGGCCCGGGCGTTACTTCCGGAGGTCCCAGGCCTTGCTGGG 2375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9510 CAGCTACACCATGCAGGAATTGGCCCGGGCGTTACTTCCGGAGGTCCCAGGCCTTGTGAGT 9451

Qy   2376 CCA 2378
          |||
Db   9450 GCA 9448
```

RESULT 8
US-09-954-456-176/c
; Sequence 176, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 176
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-176

Query Match 7.0%; Score 292; DB 10; Length 311;
Best Local Similarity 98.1%; Pred No. 7.5e-52;
Matches 306; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
Qy   3843 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTATAAACCTTCTTCTAGCAC 3902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTATAAACCTTCTTCTAGCAC 252

Qy   3903 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATTACTCATCATCAGCAGGC 3962
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATTACTCATCATCAGCAGGC 192

Qy   3963 AAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGCTGCAGAACTC 4022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGCGGCTGCAGAACTC 132

Qy   4023 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCTCATCCTGAGGAAATAAAAC 4082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCTCATCCTGAGGAAATAAAAC 72

Qy   4083 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGTTGAGGAAATAAAAC 4142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGTTGAGGAAATAAAAC 13

Qy   4143 TCCCTAAAAAAA 4154
          |||||  | |||
Db     12 TCCCTGGAGAAA 1
```

RESULT 9
US-09-954-456-861/c
; Sequence 861, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26

```
;  PRIOR APPLICATION NUMBER: US/60/246,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,700
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,940
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/236,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 861
;   LENGTH: 411
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-456-861

  Query Match             7.0%;  Score 292;  DB 10;  Length 311;
  Best Local Similarity   98.1%;  Pred. No. 7.5e-52;
  Matches  306;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy    3844 TCAGCCAGCATGGCTGGGCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 3903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 TCAGCCACATGGCTGTGCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 252

Qy    3904 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGTCACCAGCAGGC 3963
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGTCACCAGCAGGC 192

Qy    3964 AAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 4023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 132

Qy    4024 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCAC CCACCCAGCTCTCCC 4083
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCCACCCAGCTCTCCC 72

Qy    4084 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 4143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 13

Qy    4144 TCCCTAAAAAAA 4155
           |||||  | |||
Db      12 TCCCTGGAGAAA 1


RESULT 10
US-09-880-107-1520/c
; Sequence 1520, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Barnes, Darci T.
;  APPLICANT: Woolley, Joseph C.
;  APPLICANT: Scherf, Uwe
;  APPLICANT: Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5029 WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE: 2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,370
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,854
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3960
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1520
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: GenBank Accession No. US20020142981A1 AA021277

US-09-880-107-1520

  Query Match             7.0%;  Score 292;  DB 10;  Length 311;
  Best Local Similarity   98.1%;  Pred. No. 7.5e-52;
  Matches  306;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy    3844 TCAGCCAGCATGGCTGGGCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 3903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 TCAGCCACATGGCTGTGCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 252

Qy    3904 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGTCACCAGCAGGC 3963
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGTCACCAGCAGGC 192

Qy    3964 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 4023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 132

Qy    4024 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCAC CCACCCAGCTCTCCC 4083
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCCACCCAGCTCTCCC 72

Qy    4084 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 4143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 13

Qy    4144 TCCCTAAAAAAA 4155
           |||||  | |||
Db      12 TCCCTGGAGAAA 1


RESULT 11
US-09-864-761-7087/c
; Sequence 7087, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT: Rank, David R.
;  APPLICANT: Hanzel, David K.
;  APPLICANT: Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACIDS PE
;  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE: 2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,466
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/642,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,459
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7087
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
US-09-864-761-7087

Query Match 4.7%; Score 196; DB 10; Length 526;
Best Local Similarity 100.0%; Pred. No. 7.5e-32;
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1656 GCTGGCAGTTTGGCTCTGCCGGGGGCCCTTCCGGACTCTTTCCTGCCGACATAGTG 1715
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141 GCTGGCAGTTTGGCTCTGCCGGGGGCCCTTCCGGACTCTTTCCTGCCGACATAGTG 197

Qy 1716 CAGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAG 1775
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 198 CAGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAG 254

Qy 1776 GGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGG 1835
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255 GGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGGTGAGG 314

Qy 1836 AAGATGGGAGAGGGAC 1851
||||||||||||||||
Db 315 AAGATGGGAGAGGGAC 330

RESULT 12
US-09-864-761-23818/c
; Sequence 23818, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,459
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23818
; LENGTH: 170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EST_HUMAN HIT: AV649878 1, EVALUE 5.00e-91
; OTHER INFORMATION: SWISSPROT HIT: P05895, EVALUE 1.00e+00
; OTHER INFORMATION: NT HIT gi|11426247, EVALUE 1.50e+00
US-09-864-761-23818

Query Match 4.1%; Score 170; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1660 GCTGGCAGTTTGGCTCTGCCGGGGGCCCTTCCGGACTCTTTCCTGCCGACATAGTGCAGC 1719
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 170 GCTGGCAGTTTGGCTCTGCCGGGGGCCCTTCCGGACTCTTTCCTGCCGACATAGTGCAGC 111

Qy 1720 CGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTC 1779
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110 CGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTC 51

Qy 1780 AGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAG 1829
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 50 AGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAG 1

RESULT 13
US-09-864-761-24717/c
; Sequence 24717, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761

```
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/639,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/808,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24717
;   LENGTH: 233
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =    3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =    3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  2.6
;   OTHER INFORMATION: SWISSPROT HIT: P70085, EVALUE 2.9e-01
;   OTHER INFORMATION: NT HIT: AF004836.1, EVALUE 8.70e-00
;   OTHER INFORMATION: EST_HUMAN HIT: AW753961.1, EVALUE 4.00e-79
US-09-864-761-24717

  Query Match             3.6%;  Score 149.8;  DB 10;  Length 233;
  Best Local Similarity   79.1%;  Pred. No. 2.3e-22;
  Matches  178;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    3042 GCTCAGTTGGTGCTGGGCCCCTGCAGCCCGCCCAATACCTCAACTAGTGTGGTAGTTGGACCAGG 3101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     233 GCTCAGTTGGTGCTGGGCCCCTGCAGCCCGCCCAATACCTCAACTAGTGTGGTAGTTGGACCAGG 174

Qy    3102 ACCTCAGCCTGCACAGCCCGGCGGCTCCACTGGGAGACCCACTGCACTTCGATAACCTCTA 3161
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173 ACCTCAGCCTGCACAGCCCGGCGGCTCCACTGCGAGACCCACTGCACTTCGATAACCTCTA 114

Qy    3162 CCTACATCAGCACCCACTACAGCCTAGGTGCTGCTGAGTACTTCAGTGAAAGTTGTCAG 3221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     113 CCTACATCAGCACCCACTACAGCCTAGGTGCTGCTGAGTACTTCAGTGAAAGTTGTCAG 54

Qy    3222 TCAGCGCCAAGGCAGACGGCGCAGCTCGCCAGGCTGGCCCGCCTCC 3266
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      53 TTGAGGGCGCGGCTTCCTGCCTGGGCGACCTCCAAGCCCCTCTCC 9


RESULT 14
US-09-864-761-8013/c
; Sequence 8013, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER:  US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER:  US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER:  US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/639,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER:  GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER:  US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER:  PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER:  US 09/808,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER:  US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8013
;   LENGTH: 566
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION:  EXPRESSED IN LUNG, SIGNAL =    3
;   OTHER INFORMATION:  EXPRESSED IN HEART, SIGNAL = 2.5
;   OTHER INFORMATION:  EXPRESSED IN BONE MARROW, SIGNAL =    3
;   OTHER INFORMATION:  EXPRESSED IN PLACENTA, SIGNAL = 2.8
;   OTHER INFORMATION:  EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION:  EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
;   OTHER INFORMATION:  EXPRESSED IN BRAIN, SIGNAL =  2.6
US-09-864-761-8013

  Query Match             3.6%;  Score 149.8;  DB 1;  Length 566;
```

```
 Best Local Similarity   79.1%;  Pred. No. 3e-22;
 Matches  178;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    3042 GGCAGTTGGTGGGGCTTCTGAGCCCCCCGAATACCTCAACAGCGTGGTAGTGGACCAGG 3101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 GGCAGTTGGTGGGGCTTCTGAGCCCCCCGAATACCTCAACAGCGTGGTAGTGGACCAGG 320

Qy    3102 ACGTGAGGCTTCACAGCCCCACTGCACTTCGATAACTCCA 3141
           ||||||||||||||||||||||||||||||||||||||||
Db     319 ACGTGAGGCTTCACAGCCCCACTGCACTTCGATAACTCCA 280

Qy    3142 CCTACATCAGCACCCACTACAGCCAGGTGCTGTGGGACTACCTT 3185
           |||||||||||||||||||||||||||||||||||||||||||
Db     279 CCTACATCAGCACCCACTACAGCCAGGTGCTGTGGGACTACCTT 236

Qy    3186 CAGG 3189
           ||||
Db     235 CAGG 232

Qy    3190 GGAAGCTGCCAGT-CAGCGCCAAGGCAGACGGGCAGCTCCCCAGGCT--GGCCGCCCTGC 3246
           |||||||||||||  |||||||||||||||||||||||||||||||  ||  |||||||||
Db     231 GTAAGCTGCCAGTTNNAGCGCCAAGGCAGACGGGCAGCATCCCCNNAGGCTTGGCCGCCCTGC 172

Qy    3247 AGCAGCT-CAGCAAGGCCAACAGG-AATACCCCCTCAGGCAGG--AGCTGCTAGCTTA 3301
           |||||||  ||||||||||||||||  |||||||||||||||  ||||   ||  |||||
Db     171 AGCAGCTTGNAGCAAGGCCAACAGGGAATTACCCCCTCAGCNGCAGGAGCTGCTAAGNTTA 112

Qy    3302 CGTGCCAAAGCAGCTGCAACGGCAGGTG-AACAGGGCCTGCATCAAGAACC 3351
           |||||||||||||||| |||||  |||| |||||||||||||| |||| | |
Db     111 CGTGCCAAAGCAGCTGNAACGGNAGGTGTAACACGGCCTGCANCAAGAAC 62

Search completed: January 10, 2003, 02:21:30
Job time : 290.711 secs
```

```
;  LOCATION: (169)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (170)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (181)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (196)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (228)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (244)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (264)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (270)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (276)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (290)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (308)
;  OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1780

  Query Match             3.0%;  Score 125.8;  DB 10;  Length 312;
  Best Local Similarity   80.8%;  Pred. No. 2.4e-17;
  Matches  235;  Conservative    0;  Mismatches   43;  Indels   15;  Gaps    8;

Qy    3094 GGACCAGGACGTGAGCCTGCACAGCCCCCCGAATACCTCAACAGCGTGGTAGTGGACCAGG 3153
           |||||||||||| ||||| ||||||||||||||||| |||||| ||||||||||||||||||
Db       8 GGACCAGGACGTNAGCCTNCACAGCCCCCCGAATACCTCAACAGCGTGGTAGTGGACCAGG 67

Qy    3153 ATAACTCCACCTACATCAGCACCCACTACAGCCAGGTGCTGTGGGACTACCTT--CAGG 3209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db      68 ATAACTCCACCTACATCAGCACCCACTACAGCCAGGTGCTGTGGGAACTACCTTGNAGGG 127

Qy    3210 GGAAGCTGCCAGT-CAGCGCCAAGGCAGACGGGCAGCTCCCCAGGCT--GGCCGCCCTGC 3266
           |||||||||||||  |||||||||||||||||||||||||||||||  ||  |||||||||
Db     128 GTAAGCTGCCAGTNNAGCGCCAAGGCAGACGGGCAGCATCCCCNNAGGCTTGGCCGCCCTGC 187

Qy    3267 AGCAGCT-CAGCAAGGCCAACAGG-AATACCCCCTCAGGCAGG--AGCTGCTAGCTTA 3321
           |||||||  ||||||||||||||||  |||||||||||||||  ||||   ||  |||||
Db     188 AGCAGCTTGNAGCAAGGCCAACAGGGAATTACCCCCTCAGCNGCAGGAGCTGCTAAGNTTA 247

Qy    3322 CGTGCCAAAGCAGCTGCAACGGCAGGTG-AACAGGGCCTGCATCAAGAACC 3371
           |||||||||||||||| |||||  |||| |||||||||||||| |||| | |
Db     248 CGTGCCAAAGCAGCTGNAACGGNAGGTGTAACACGGCCTGCANCAAGAACC 298

Search completed: January 10, 2003, 02:21:30
Job time : 290.711 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

M protein - protein search, using sw model

Run on: January 9, 2003, 07:23:38 ; Search time 54.0375 Seconds
(without alignments)
1938.191 Million cell updates/sec

Title: US-09-803-126-6
Perfect score: 4129
Sequence: 1 MAVSKRGIVPVFVGIPPPY ... WPPCSTSARPTGIPPQGRTC 786

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133259620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_121002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4129	100.0	786	22	AAU07406	Human oestrogen-re
2	2217.5	53.7	437	22	AAU07400	Human oestrogen-re
3	1657.5	40.1	337	22	AAU87214	Novel central nerv
4	1655.5	40.1	333	22	AAU87523	Novel central nerv
5	1655.5	40.1	333	22	AAU17193	Novel signal trans
6	1648.5	39.9	470	21	AAY57895	Human transmembran
7	1544	37.4	434	22	AAU20565	Human secreted pro
8	1537	37.2	1337	22	AAU07399	Mouse oestrogen-re
9	416	10.1	2167	22	ABB60369	Drosophila melanog
10	344.5	8.3	563	22	ABB58859	Drosophila melanog
11	306	7.4	56	22	ABB41520	Peptide #9026 enco
12	306	7.4	56	22	ABB25387	Protein #7386 enco
13	306	7.4	56	22	AAM62390	Human brain expres
14	306	7.4	56	22	AAM75200	Human bone marrow
15	306	7.4	56	22	AAM25312	Peptide #9349 enco
16	306	7.4	56	23	ABG44860	Human peptide enco
17	304.5	7.4	2424	22	ABB58924	Drosophila melanog
18	278.5	6.7	2329	22	ABB62828	Drosophila melanog
19	231.5	5.6	1651	22	ABG14648	Novel human diagno
20	198.5	4.8	1269	21	AAY72955	A. thaliana enviro
21	194.5	4.7	673	22	ABG09931	Novel human diagno
22	163.5	4.0	343	22	ABG14647	Novel signal trans
23	153.5	3.7	1031	22	AAU17082	Human protein SEQ
24	153.5	3.7	1452	22	AAM78778	Novel human diagno
25	153.5	3.7	1515	22	ABG10624	Angiogenesis assoc
26	153.5	3.7	2048	22	AAE11991	Angiogenesis assoc
27	153.5	3.7	2057	22	AAE11890	Novel human protei
28	153.5	3.7	2058	23	ABB97219	Novel human diagno
29	153.5	3.7	2408	22	ABG10631	Human polypeptide,
30	151.5	3.7	984	22	AAM93762	Human protein SEQ
31	150	3.6	1469	22	AAM79762	Drosophila melanog
32	147.5	3.6	1309	22	ABB59882	Drosophila melanog
33	147.5	3.6	3536	22	ABB65480	Novel human diagno
34	145.5	3.5	921	22	ABG20416	Novel human diagno
35	145	3.5	1236	22	ABG08165	Human Scarl protei
36	143	3.5	559	22	AAM52317	Amino acid sequenc
37	143	3.5	559	22	AAG67345	Human Pablo (pro-a
38	143	3.5	559	22	AAP20494	Human WAVE1 (WAVE/
39	143	3.5	559	22	AAP73399	Human NNG2 (long f
40	143	3.5	709	21	AAB19691	Novel human diagno
41	142.5	3.5	713	22	ABG10632	Novel human protei
42	142.5	3.5	1215	23	ABB97489	Novel human diagno
43	141	3.4	1321	22	ABG25416	Human Alfa [H sp
44	140.5	3.4	216	22	AAB80892	Human [H [sin sequ
45	140.5	3.4	282	22	AAB95859	

ALIGNMENTS

RESULT 1
AAU07406
ID AAU07406 standard; Protein; 786 AA.
XX
AC AAU07406;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human oestrogen-regulated unconventional myosin-related protein, MRP #2.
XX
KW Human; oestrogen-regulated unconventional myosin-related protein, MRP;
KW auditory; antiatherosclerotic; osteopathic, neuroprotective, nootropic;
KW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
KW Alzheimer's disease, Parkinson's disease; chromosome 17q, immunomodulatory.
XX
OS Homo sapiens.
XX
PN WO200168866 A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Brooks AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS13823.

```
AA
PT   Novel oestrogen regulated unconventional myosin-related protein and
PT   polynucleotide for modulating effects of oestrogen in mammalian cells to
PT   treat, prevent atherosclerosis, breast cancer and osteoporosis
AA
PS   Claim 1; Fig 6; 108pp; English.
XX
CC   The invention relates to an isolated oestrogen-regulated unconventional
CC   myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC   effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC   mammalian cell. When the level of expression or activity of MRP is
CC   increased, the development of atherosclerosis or osteoporosis is
CC   inhibited and when the level is decreased the development of breast
CC   cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC   useful for detecting the presence of oestrogen signaling in a mammalian
CC   cell, by detecting the expression or activity of MRP or the level of MRP
CC   mRNA in the cell. The presence of oestrogen signaling is used to
CC   determine the responsiveness of the cell to the oestrogen or the tissue-
CC   specific distribution of oestrogen signaling in a mammal. MRP is useful
CC   as a marker gene for the investigation of tissue-specific and oestrogen
CC   receptor-specific ligands such as oestrogen receptor antagonists and
CC   agonists. Modulators, recombinant forms, derivatives, variants or
CC   fragments of MRP are useful for enhancing or inhibiting oestrogen
CC   signaling in cells and for treatment of breast cancer or other cancers
CC   and for preventing or treating wide variety of oestrogen-related
CC   diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC   disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC   preventing and/or treating hearing loss and deafness. MRP polypeptides
CC   are useful for generating monoclonal and polyclonal antibodies useful for
CC   identifying oestrogen responsive cells, particularly cells responsive to
CC   oestrogen through an oestrogen receptor alpha or beta, to identify
CC   alleles, interspecies homologues and polymorphic variants and in
CC   immunoassays to detect MRP. The present sequence represents the amino
CC   acid sequence of human MRP variant #2 as described in the invention.
XX
SQ   Sequence   786 AA,

  Query Match        100.0%;  Score 4129;  DB 22;  Length 786;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVLVPFSDVVYVFVQLDRFPAFLFPIDPEPALAFLGINDAHESFPMLLPSPGRGEPPAV 60
Db       1 MVLVPFSDVVYVFVQLDRFPAFLFPIDPEPALAFLGINDAHESFPMLLPSPGRGEPPAV 60

Qy      61 APDPFAPLQLVPPSSIPEPVVPPIIPPPQPFLPIAHTPPPPAPPLPLPFPPQTLSAPPPP 120
Db      61 APDPFAPLQLVPPSSIPEPVVPPIIPPPQPFLPIAHTPPPPAPPLPLPLPPQTLSAPPPP 120

Qy     121 LTQPVPDVQPVSTQLLAPSQSVPPSYTGTPWKLPLPPFVPYPPENFSHPYYIPLLCFQILP 180
Db     121 LTQPVPDVQPVSTQLLAPSQSVPPSYTGTPWKLPLPPFVPYPPENFSHPYYIPLLCFQILP 180

Qy     181 PTPPPPPPPRIQQNEPPPMPLLQPLEVLLPLTTTPPPVPPRIVVAAPPNWANYPPPPP 240
Db     181 PTPPPPPPPRIQQNEPPPMPLLQPLEVLLPLTTTPPPVPPRIVVAAPPNWANYPPPPP 240

Qy     241 VSPPSSPDVQLLAVSHPGLPLLEVTQPGLPPPQLPPLCSYSPAEVLGVECRGGSTLELS 300
Db     241 VSPPSSPDVQLLAVSHPGLPLLEVTQPGLPPPQLPPLCSYSPAEVLGVECRGGSTLELS 300

Qy     301 LPPPQLNLHTAPAPAIEALVPLPLNEIPPPQPPYIALPCYITTNPSLLSPPRPQLIPLLP 360
Db     301 LPPPQLNLHTAPAPAIEALVPLPLNEIPPPQPPYIALPCYITTNPSLLSPPRPQLIPLLP 360

Qy     361 VATLDLGWQPLAGGPSVLPLPALDVLIAAAAPLPPPPPLPPPTWHPGLSLNTEPDLAPWSP 420
Db     361 VATLDLGWQPLAGGPSVLPLPALDVLIAAAAPLPPPPPLPPPTWHPGLSLNTEPDLAPWSP 420

Qy     421 ASPRPAHPWSQAHSDDSPATSLSSVASAPLPDSHSYTMLEPAPPYPPPSQALLGQTDGGA 480
Db     421 ASPRPAHPWSQAHSDDSPATSLSSVASAPLPDSHSYTMLEPAPPYPPPSQALLGQTDGGA 480

Qy     481 AGPGTPLDLVQYTFAPSGLALLQLLSSTEVRLAVAGTLALMPKMPEMKLDVPLPMELVPEL 540
Db     481 AGPGTPLDLVQYTFAPSGLALLQLLSSTEVRLAVAGTLALMPKMPEMKLDVPLPMELVPEL 540

Qy     541 LPLCQEPFLPDRIYCQVIEQVTGHPPPPH-TRGWSPLSLGLMPYLTRPLQP 590
Db     541 LPLCQEPFLPDRIYCQVIEQVTGHPPPPHPTRGWSPLSLGLMPYLTRPLQP 592

Qy     591 QSHLPLAPSQLPHLPTVPTSSPPMPPGPMPAGLPAAPLPLPLPPGVDRTNIQ 642
Db     593 PSPPQPLAPSQLPHLPTVPTSSPPMPPGPMPAGLPAAPLPLPLPPGVDRTNIQ 644

Qy     643 TPIVAAEVQEELCRQMSIIEPQFVQEFALPLIPEPPQLVPPLQPAEYINPVVVDQLVSLH 702
Db     645 TPTVAAEVQEELCRQMRITPPQFVQEFALPLIPEPPQLVPPLQPAEYINPVVVDQLVSLH 704

Qy     703 SGSGTGRPHQTCITIPLALFTAPLSSTFPGSCQSAPPQTRNSPSWIPCSTPARPTGII 760
Db     705 SGSGTGRPHQTCITIPLALFTAPLSSTFPGSCQSAPPQTRNSPSWIPCSTPARPTGII 762

Qy     761 PQGPTC 786
Db     763 PQGPTC 788


RESULT 2
AAW07400
ID   AAW07400 standard; protein; 437 AA.
XX
AC   AAW07400;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human oestrogen-regulated unconventional myosin related protein (MRP) #1.
XX
KW   Human; oestrogen regulated unconventional myosin related protein; MRP;
KW   auditory; antiatherosclerotic; osteopathic; neuroprotective; neurotropic;
KW   antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW   osteoporosis, breast cancer, cardiovascular disease; deafness;
KW   Alzheimer's disease, Parkinson's disease; chromosome 17q25; immunoassay.
XX
OS   Homo sapiens.
XX
RN   [1]
RP   20-SEP-2001.
XX
RF   12-MAR-2002; 2002WO-US08660.
XX
PR   16-MAR-2000; 2000US-188488P.
PR   09-MAR-2001; 2001US-0803126.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Brooks AR,  Deng GG,  Rubanyi GM;
XX
DR   WPI; 2002-602747/68.
DR   N-PSDB; AAZ13821.
XX
PT   Novel oestrogen regulated unconventional myosin related protein and
PT   polynucleotide for modulating effects of oestrogen in mammalian cells to
PT   treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS   Claim 1; Fig 4; 108pp; English.
XX
CC   The invention relates to an isolated oestrogen regulated unconventional
CC   myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC   effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC   mammalian cell. When the level of expression or activity of MRP is
CC   increased, the development of atherosclerosis or osteoporosis is
CC   inhibited and when the level is decreased the development of breast
CC   cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC   useful for detecting the presence of oestrogen signaling in a mammalian
```

CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen-related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the amino
CC acid sequence of human MRP variant #1 as described in the invention.
XX
SQ Sequence 437 AA;

Query Match 53.7%; Score 2217.5; DB 22; Length 437;
Best Local Similarity 97.5%; Pred. No. 2.2e-186;
Matches 426; Conservative 0; Mismatches 0; Indels 11; Gaps 1;

```
Qy       1 MYQSRPGPVPVPVQPSRPPKAPLPKIDPKDPAIAKLGINGAHSSPPMLSPSPGKGPPPAV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MYQSRPGPVPVPVQPSRPPKAPLPKIDPKDPAIAKLGINGAHSSPPMLSPSPGKGPPPAV 60

Qy      61 APPPKAPLQLGPSSSIKPKQGPLLDLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAEPPC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 APPPKAPLQLGPSSSIKPKQGPLLDLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAEPPC 120

Qy     121 LTQPVPDQGVSTQLLAPSGSVCPSYTGTPWPLPLPKEVPYPPPNPSHPYYLPLLCEQILP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LTQPVPDQGVSTQLLAPSGSVCPSYTGTPWPLPLPKEVPYPPPNPSHPYYLPLLCEQILP 180

Qy     181 DTPSESCIPIPQNEPPPKMKDLLGGLEVDLDSLTTTEDSVKKPIVVAARDNWANYPSRPPP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DTPSESCIPIPQNEPPPKMKDLLGGLEVDLDSLTTTEDSVKKRIVVAARDNWANYPSRPPP 240

Qy     241 VSGESGSDVQDLAVSHRGLPLLKVTQGPGLPPDQLKILCSYSFAEVLGVECPGGSTLELS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VSGESGSDVQDLAVSHRGLPLLKVTQGPGLPPDQLKILCSYSFAEVLGVECPGGSTLELS 300

Qy     301 LPSEQIVIHTAPAPAIPALVELPLNELKKDSGYVIALPSYITDNCSLLSPHPGDLIKLLP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 LPSEQIVIHTAPAPAIPALVELPLNELKKDSGYVIALPSYITDNCSLLSPHPGDLIKLLP 360

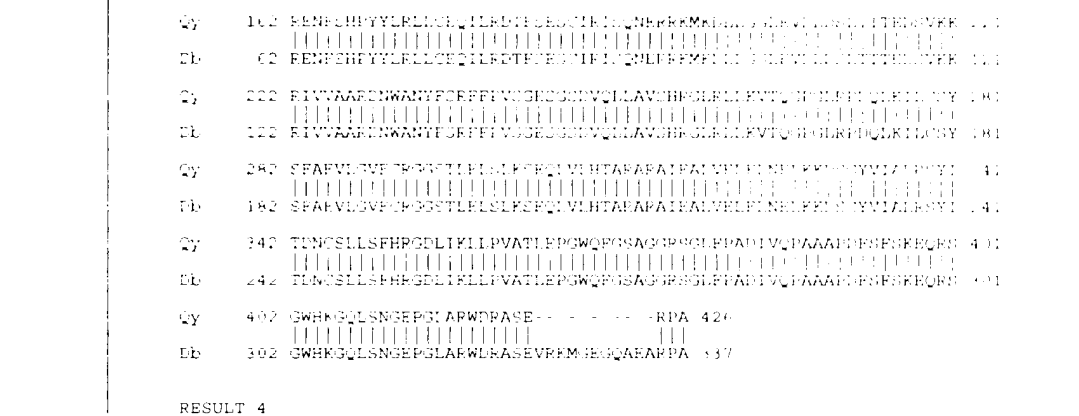
Qy     361 VATLEPGWQPGSAQGRSGLPPADIVQPAAAPDPSPSKEQPSQGWHKGQLSNGEPGLARWDR 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 VATLEPGWQPGSAQGRSGLPPADIVQPAAAPDPSPSKEQPSQGWHKGQLSNGEPGLARWDR 420

Qy     421 ASE----------RPA 426
           |||          |||
Db     421 ASEVRKMGEGQAEARPA 437
```

RESULT 3
AAU87214
ID AAU87214 standard; Protein; 337 AA.
XX
AC AAU87214;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #124.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001
XX
PF 17-JAN-2001; 2001WO-US01332
XX
PR 31-JAN-2000; 2000US-0179065
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135
PR 07-JUL-2000; 2000US-0216647
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496
PR 14-JUL-2000; 2000US-0218290
PR 26-JUL-2000; 2000US-0220963
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266
PR 14-AUG-2000; 2000US-0225267
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868
PR 22-AUG-2000; 2000US-0227182
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924
PR 01-SEP-2000; 2000US-0229287
PR 01-SEP-2000; 2000US-0229343
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234447.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 28-SEP-2000; 2000US-0236327.
PR 28-SEP-2000; 2000US-0236367.
PR 28-SEP-2000; 2000US-0236368.
PR 28-SEP-2000; 2000US-0236369.
PR 28-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 19-OCT-2000; 2000US-0239935.
PR 19-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240060.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246612.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249298.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250189.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251630.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR N-PSDB; ABK43544.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 9; SEQ ID No 732; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and (...)peptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 40.1%; Score 1657.5; DB 27; Length 837;
Best Local Similarity 96.1%; Pred. No. 3.2e-197;
Matches 823; Conservative 0; Mismatches 2; Indels 11; Gaps 1;



RESULT 4

```
AU87523
D   AAU87523 standard; Protein; 333 AA.
X
C   AAU87523;
X
T   05-JUN-2002  (first entry)
X
E   Novel central nervous system protein #433.
X
W   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
W   hyperproliferative disorder; neoplasm, cardiovascular disorder;
W   cardiac arrest; cerebrovascular disorder, ischaemia; angiogenesis;
W   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
W   acquired immunodeficiency virus, dysphagia; gastrointestinal disorder;
W   adenocarcinoma; reproductive system disorder, testicular feminisation;
W   endocrine disorder; diabetes; cancer, leukaemia; neovascularisation;
W   respiratory disorder; renal disorder, kidney failure; blood disorder;
W   myocardial infarction; wound healing, cell proliferation, skin aging;
W   food additive; food preservative; gene therapy.
X
S   Homo sapiens.
X
N   WO200155318-A2.
X
D   02-AUG-2001
X
F   17-JAN 2001; 2001WO-US01143.
X
R   31-JAN-2000; 2000US-0179065.
R   04-FEB-2000; 2000US-0180124.
R   24-FEB-2000; 2000US-0184664.
R   02-MAR-2000; 2000US-0186360.
R   16-MAR-2000; 2000US-0189874.
R   17-MAR-2000; 2000US-0190076.
R   18-APR-2000; 2000US-0198123.
R   19-MAY-2000; 2000US-0205515.
R   07-JUN-2000; 2000US-0209467.
R   28-JUN-2000; 2000US-0214886.
R   30-JUN-2000; 2000US-0215135.
R   07-JUL-2000; 2000US-0216647.
R   07-JUL-2000; 2000US-0216880.
R   11-JUL-2000; 2000US-0217487.
R   11-JUL-2000; 2000US-0217490.
R   14-JUL-2000; 2000US-0218290.
R   26-JUL-2000; 2000US-0220963.
R   26-JUL-2000; 2000US-0220964.
R   14-AUG-2000; 2000US-0224518.
R   14-AUG-2000; 2000US-0224519.
R   14-AUG-2000; 2000US-0225213.
R   14-AUG-2000; 2000US-0225214.
R   14-AUG-2000; 2000US-0225266.
R   14-AUG-2000; 2000US-0225267.
R   14-AUG-2000; 2000US-0225268.
R   14-AUG-2000; 2000US-0225270.
R   14-AUG-2000; 2000US-0225447.
R   14-AUG-2000; 2000US-0225757.
R   14-AUG-2000; 2000US-0225758.
R   14-AUG-2000; 2000US-0225759.
R   18-AUG-2000; 2000US-0226475.
R   22-AUG-2000; 2000US-0226681.
R   22-AUG-2000; 2000US-0226868.
R   22-AUG-2000; 2000US-0227182.
R   23-AUG-2000; 2000US-0227609.
R   30-AUG-2000; 2000US-0228324.
R   01-SEP-2000; 2000US-0229287.
R   01-SEP-2000; 2000US-0229342.
R   01-SEP-2000; 2000US-0229344.
R   01-SEP-2000; 2000US-0229345.
R   05-SEP-2000; 2000US-0229509.
R   05-SEP-2000; 2000US-0229513.
R   06-SEP-2000; 2000US-0230437.
R   06-SEP-2000; 2000US-0230438.
R   08-SEP-2000; 2000US-0231242.
R   08-SEP-2000; 2000US-0231243.
R   08-SEP-2000; 2000US-0231244.
R   08-SEP-2000; 2000US-0231413.
R   08-SEP-2000; 2000US-0231414.
R   08-SEP-2000; 2000US-0232080.
R   08-SEP-2000; 2000US-0232081.
R   12-SEP-2000; 2000US-0231968.
R   14-SEP-2000; 2000US-0232397.
R   14-SEP-2000; 2000US-0232398.
R   14-SEP-2000; 2000US-0232399.
R   14-SEP-2000; 2000US-0232400.
R   14-SEP-2000; 2000US-0232401.
R   14-SEP-2000; 2000US-0233063.
R   14-SEP-2000; 2000US-0233064.
R   14-SEP-2000; 2000US-0233065.
R   21-SEP-2000; 2000US-0234223.
R   21-SEP-2000; 2000US-0234274.
R   25-SEP-2000; 2000US-0234997.
R   25-SEP-2000; 2000US-0234998.
R   26-SEP-2000; 2000US-0235484.
R   27-SEP-2000; 2000US-0235834.
R   27-SEP-2000; 2000US-0235836.
R   29-SEP-2000; 2000US-0236327.
R   29-SEP-2000; 2000US-0236367.
R   29-SEP-2000; 2000US-0236368.
R   29-SEP-2000; 2000US-0236369.
R   29-SEP-2000; 2000US-0236370.
R   02-OCT-2000; 2000US-0236802.
R   02-OCT-2000; 2000US-0237037.
R   02-OCT-2000; 2000US-0237038.
R   02-OCT-2000; 2000US-0237039.
R   02-OCT-2000; 2000US-0237040.
R   13-OCT-2000; 2000US-0239935.
R   13-OCT-2000; 2000US-0239937.
R   20-OCT-2000; 2000US-0240960.
R   20-OCT-2000; 2000US-0241221.
R   20-OCT-2000; 2000US-0241785.
R   20-OCT-2000; 2000US-0241786.
R   20-OCT-2000; 2000US-0241787.
R   20-OCT-2000; 2000US-0241808.
R   20-OCT-2000; 2000US-0241809.
R   20-OCT-2000; 2000US-0241826.
R   01-NOV-2000; 2000US-0244617.
R   08-NOV-2000; 2000US-0246474.
R   08-NOV-2000; 2000US-0246475.
R   08-NOV-2000; 2000US-0246476.
R   08-NOV-2000; 2000US-0246477.
R   08-NOV-2000; 2000US-0246478.
R   08-NOV-2000; 2000US-0246523.
R   08-NOV-2000; 2000US-0246524.
R   08-NOV-2000; 2000US-0246525.
R   08-NOV-2000; 2000US-0246526.
R   08-NOV-2000; 2000US-0246527.
R   08-NOV-2000; 2000US-0246528.
R   08-NOV-2000; 2000US-0246532.
R   08-NOV-2000; 2000US-0246609.
R   08-NOV-2000; 2000US-0246610.
R   08-NOV-2000; 2000US-0246611.
R   08-NOV-2000; 2000US-0246613.
R   17-NOV-2000; 2000US-0249207.
R   17-NOV-2000; 2000US-0249208.
R   17-NOV-2000; 2000US-0249209.
R   17-NOV-2000; 2000US-0249210.
R   17-NOV-2000; 2000US-0249211.
R   17-NOV-2000; 2000US-0249212.
R   17-NOV-2000; 2000US-0249213.
R   17-NOV-2000; 2000US-0249214.
R   17-NOV-2000; 2000US-0249215.
R   17-NOV-2000; 2000US-0249216.
R   17-NOV-2000; 2000US-0249217.
R   17-NOV-2000; 2000US-0249218.
R   17-NOV-2000; 2000US-0249244.
R   17-NOV-2000; 2000US-0249245.
```

PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 21-DEC-2000; 2000US-0258160.
PR 21-DEC-2000; 2000US-0258391.
PR 05-DEC-2000; 2000US-0251019.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR N-PSDB; AAR43858.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives
XX
PS Claim 1; SEQ ID NO 1041; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC [illegible] and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as food additives or preservatives to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 40.1%; Score 1650.5; DB 22; Length 333;
Best Local Similarity 96.7%; Pred. No. 4.8e-137;
Matches 322; Conservative 0; Mismatches 9; Indels 11; Gaps 1;

Qy 139 LPLPETPSGTLSAEPPCLTQPVPDQGVSTQLIAPSGSVCFSYTGTPWKLFLRKEVPYPPEN 184
Db 1 LPLPETPSGTLSAEPPCLTQPVPDQGVSTQLIAPSGSVCFSYTGTPWKLFLRKEVPYPPEN 60

Qy 185 PSHDYYIPIICPQTIKPQTIKPTPAPSTIKTQNEPPPMPGSLDSLEVSLISLTTTPDSVPPPIV 244
Db 61 PSHDYYIPIICPQTIKPQTIKPSPSTIKTQNEPPPMPGSLDSLEVSLISLTTTPDSVPPPIV 120

Qy 245 VAAPDNWAVPPRPPPVSRPGQTVQLLAVSHPGRRPPLLPFLVQTGRSLPPDLPILVYVPA 304
Db 121 VAAPDNWAVPPRPPPVSRPGQTVQLLAVSHPGRRPPLLPFLVQTGRSLPPDLPILVYVPA 180

Qy 305 PPI MPGPGQSTLPIQIPLSIVSIRARAAIPAIVRLPINPLPPISDVVALSPYITKN 360
Db 181 PVIGVPPPGQSTLPIQIPLSIVSIRARAAIPAIVRLPINPLPPISDVVALSPYITKN 240

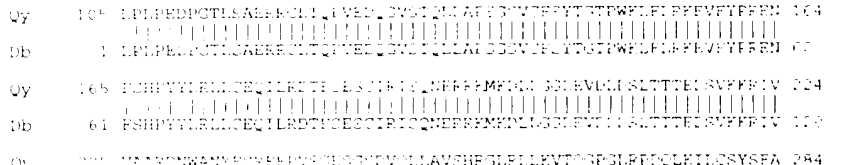
Qy 345 QSLISPHPGPLIPILIPVATIEPSWLPISAVQPVVSLPPADIVLIAAAPLPPPPEQPPSWH 404
Db 241 QSLISPHPGPLIPILIPVATIEPSWLPISAVQPVVSLPPADIVLIAAAPLPPPPEQPPSWH 300

Qy 405 FGQLSNGFPGLARWDRASE 425
Db 301 FGQISNGFPGLAPWERASEFVPEPMLEGQAEARPA 333

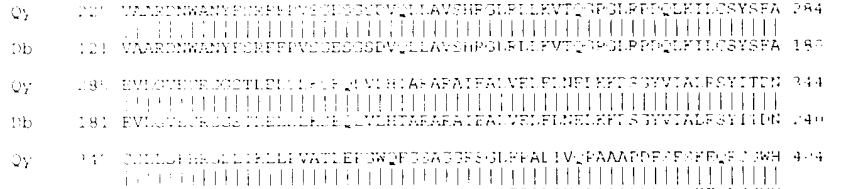
RESULT 5
AAU17193
ID AAU17193 standard; Protein; 333 AA.
XX
AC AAU17193;
XX
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID [illegible].
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; [illegible];
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection, infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular, respiratory, wound healing, [illegible], Addison's disease;
KW reproductive system, gastrointestinal; liver disorder; AIDS;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179265.
PR 04-FEB-2000; 2000US-0180628.
PR 14-FEB-2000; 2000US-0182664.
PR 02-MAR-2000; 2000US-0186360.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 10-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214986.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226869.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227900.
PR 30-AUG-2000; 2000US-0229924.
PR 01-SEP-2000; 2000US-0229587.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239837.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR N-PSDB; AAS27110.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID No 758; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 40.1%; Score 1655.5; DB 22; Length 333;
Best Local Similarity 96.7%; Pred. No. 4.8e-137;
Matches 322; Conservative 0; Mismatches 0; Indels 11; Gaps 1;



RESULT 6
AAY57877
ID AAY57877 standard; Protein; 470 AA.
XX
AC AAY57877;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN-19.
XX
KW Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011904.
XX
PR 29-MAY-1998; 98US-0087260.
PR 02-JUL-1998; 98US-0091674.
PR 22-OCT-1998; 98US-0102454.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Rumbaut R, Patterson C, Drmanac RA, Kaiser MP, Baughn MR;
PI Au Young J;
XX
DR WPI; 2000-072605/06.
DR N-PSDB; AAU56776.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders
XX
PS Claim 1; Page 125-126; 289pp; English.
XX
CC AAY56698 to AAU56776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN 1 to HTMPN 79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or
CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC deficiency or excess of HTMPN.
XX
SQ Sequence 470 AA;



RESULT 7
AAU20565
ID AAU20565 standard; Protein; 434 AA.
XX
AC AAU20565;
XX
DT 06-DEC-2001 (first entry)
XX
DE Human secreted protein, Seq ID No 557.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; arthritis;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.
XX
OS Homo sapiens.
XX
PN WO200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01347.
XX

31-JAN 2000, 2000US-0170065.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-451931/48.
N-PSDB; AAS33274.

New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions -

Claim 11; SPC ID No 562, 772pp, English.

The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, rats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The disorders include for example immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/ cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAU02342 AAU20666 represent human secreted protein amino acid sequences, and related sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences

Sequence 434 AA;

Query Match 37.4%; Score 1644; DB 22; Length 434;
Best Local Similarity 99.7%; Pred. No. 4.9e-127;
Matches 299; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy  422 SPPPAHPWSQAHSDPSPATSLSSVAYAFLPDSHSYIMQPPAPKYPPPSQALLGQTPGGAA 481
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8 SQPPAHPWSQAHSDTSPATSLSSVAYAFLPDSHSYIMQPPAPPYPPPSQALLGQTPGGAA 67

Qy  482 QKDTDSLVQYTPAPIQPSLLSLSGDVSPLAVASPLALMPPMGDQSPPPGKDEMDLLYFLL 541
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 QKDTDSLVQYTPAPIQPSLLSLSGDVSPLAVASPLALMPPMGDQSPPPGKDEMDLLYFLL 127

Qy  542 PLCQQPPLPPEIYPQVIPQVTQHPPPEHPTPGWSPLSLLTQPPPPSTPLMPYLTKPLQDS 601
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 PLCQQPPLPPEIYPQVIPQVTQHPPPEHPTPGWSPLSLLTQPPPPSTPLMPYLTKPLQDS 187

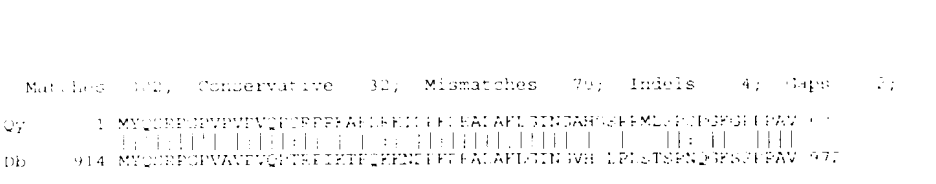
Qy  602 GPSQFLARSSQFHLQPTVKYGGPPPMPPPGPMKAFIPQQATPLLLIHLPGGVDYPTNIQT 661
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  188 GPSQFLARSSQFHLQPTVKYGGPPPMPPPGPMKAFIPQQATPLLLIHLPGGVDYPTNIQT 247

Qy  662 PTVAAEVQEELQRQMSITEPQFVLPFALPLIPKEKSQLVRPLQFAEYLNSVVVDQDVSLHS 721
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  248 PTVAAEVQEELQRQMSITEPQFVLPFALPLIPKEKSQLVRPLQFAEYLNSVVVDQDVSLHS 307
```

RESULT 8
AAU07399
ID AAU07399 standard; Protein; 1337 AA.
XX
AC AAU07399;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mouse oestrogen-regulated unconventional myosin related protein, MRP.
XX
KW Mouse; oestrogen-regulated unconventional myosin-related protein; MRP;
KW auditory; antiatherosclerotic, osteopathic, neuroprotective; nootropic;
KW antiparkinsonian, cytostatic, oestrogen receptor, atherosclerosis;
KW osteoporosis; breast cancer, cardiovascular disease, deafness;
KW Alzheimer's disease; Parkinson's disease; chromosome 11; immunogenic.
XX
OS Mus sp.
XX
PN WO200169966-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Bleicher AF, Deng CG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS12813, AAS13820.
XX
PT Novel oestrogen-regulated unconventional myosin-related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 1; Fig 1; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is useful to
CC determine the responsiveness of the cell to the oestrogen or the tissue
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor-specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer and other cancers
CC and for preventing or treating wide variety of oestrogen-related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the amino
CC acid sequence of mouse MRP as described in the invention.
XX
SQ Sequence 1337 AA;

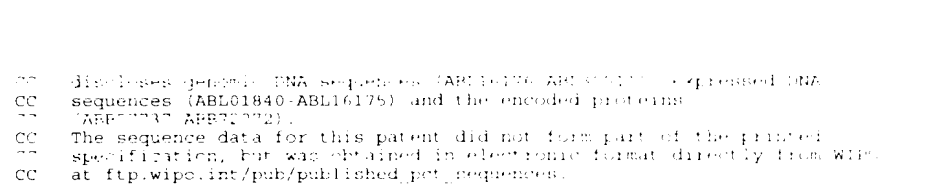
Query Match 37.2%; Score 1637; DB 22; Length 1337;
Best Local Similarity 74.0%; Pred. No. 1.1e-125;

Matches 1032; Conservative 32; Mismatches 79; Indels 4; Gaps 3;



RESULT 9
ABB62369
ID ABB62369 standard; Protein; 2167 AA.
XX
AC ABB62369;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7869
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 26-MAR-2001; 2001WO-US09231.
XX
PR 24-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (CELE-) CELERA GENOMICS CORP.
XX
PI Venter JC; Adams M; Li PWD; Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04472.
XX
PT Newly isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell cell
PT interactions
XX
PS Disclosure; SEQ ID NO 7869; 21pp; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB62192-ABB77272).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2167 AA;

Query Match 10.1%; Score 416; DB 22; Length 2167;
Best Local Similarity 23.8%; Pred. No. 1.4e-16;
Matches 183; Conservative 115; Mismatches 284; Indels 230; Gaps 27;



RESULT 10
ABB58859
ID ABB58859 standard; Protein; 561 AA.

ABB58859;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 3369.

Drosophila, developmental biology, cell signalling; insecticide; pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.
11-JUL-2000; 2000US-0614150.

(PEKE) PE COPP NY

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.
N-PSDB; ABL02962.

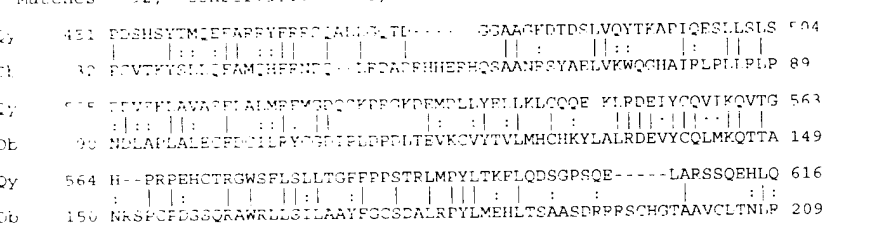
New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 3369; 21pp + Sequence Listing; English

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

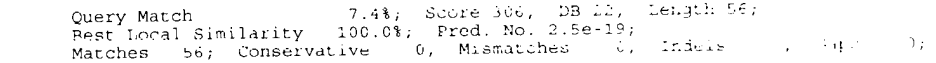
Sequence 563 AA;

Query Match 8.3%; Score 344.5; DB 22; Length 563;
Best Local Similarity 31.0%; Pred. No. 3.6e-21;
Matches 92; Conservative 61; Mismatches 135; Indels 19; Gaps 6;



RESULT 11
ABB41520
ID ABB41520 standard; Peptide; 56 AA.
XX
AC ABB41520;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #9026 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 34155; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC tetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 56 AA;

Query Match 7.4%; Score 306; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 12
ABB25387
ID ABB25387 standard; Protein; 56 AA.
XX
AC ABB25387;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #7386 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension, cardiac arrhythmia;
KW congenital heart disease.

XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30 JAN 2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analysing gene expression in human
PT hearts
XX
PS Claim 15; SEQ ID No 27157; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA41320, ABA41395). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 56 AA;

Query Match 7.4%; Score 406; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 369 WQPRSAGPRSHIPPADIVQFAAAAIFSPSFPQPSGWHPSQLSNSPFPSLARWPPASE 424
Db 1 WQPRSAGPRSHIPPADIVQFAAAAIFSPSFPQPSGWHPSQLSNSPFPSLARWPPASE 56

RESULT 13
AAM62490
ID AAM62490 standard; Protein; 56 AA.
XX
AC AAM62490;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO 44445
XX
KW Human; brain expressed; single exon; gene expression; analysis; probe;
KW microarray, Alzheimer's disease, multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.

XX
PP 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Single exon nucleic acid probes for analysing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 34495; 610pp + Sequence Listing; English.
XX
CC The present invention provides a number of novel single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 56 AA;

Query Match 7.4%; Score 406; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 369 WQPRSAGPRSHIPPADIVQFAAAAIFSPSFPQPSGWHPSQLSNSPFPSLARWPPASE 424
Db 1 WQPRSAGPRSHIPPADIVQFAAAAIFSPSFPQPSGWHPSQLSNSPFPSLARWPPASE 56

RESULT 14
AAM75200
ID AAM75200 standard; Protein; 56 AA.
XX
AC AAM75200;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human brain expressed probe encoded protein SEQ ID NO: 444 6.
XX
KW Human, brain; single expressed exon; gene expression; analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30 JAN 2001; 2001WO-US00668.
XX
PP 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.

Human genome derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow -

Example 4; SEQ ID NO. 35506, 658pp + Sequence Listing; English

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention.

Sequence 56 AA;

Query Match 7.4%; Score 306; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 368 WQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAPWDPASE 423
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 WQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAPWDPASE 56

RESULT 15
AAM35312
ID AAM35312 standard; Protein; 56 AA.
XX
AC AAM35312;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #0349 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray, human, placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 35581; 654pp, English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 56 AA;

Query Match 7.4%; Score 306; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 368 WQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAPWDPASE 423
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 WQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHKGQLSNGEPGLAPWDPASE 56

Search completed: January 10, 2003, 02:17:53
Job time : 59.0375 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein protein search, using sw model

Run on. January 10, 2003, 02:17:59 ; Search time 27.9398 Seconds
(without alignments)
2704.441 Million cell updates/sec

Title: US-09-803-126-6
Perfect score: 4129
Sequence: 1 MYQSRPGPVDVPVQPSRPPPK . . . . WPPCSTSARPTGIPPQGRTC 786

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1159.5	28.1	3511	2	A59295	unconventional myo
2	1155.5	28.0	3530	2	A59266	unconventional myo
3	688	16.7	1783	2	T42386	unconventional myo
4	545	13.2	2810	2	T22248	hypothetical prote
5	413	10.0	2175	2	A59255	myosin VIIa, long
6	409	9.9	2215	2	T30870	myosin VIIa - mous
7	363.5	8.8	2098	2	T25888	hypothetical prote
8	279.5	6.8	2121	2	A59233	myosin VII-like pr
9	273	6.6	2257	2	A59249	class VII unconven
10	208.5	5.0	1265	2	T03792	kinesin-related pr
11	203	4.9	1265	2	T07397	kinesin heavy chai
12	192	4.7	1576	1	A23662	myosin I, high mol
13	167.5	4.1	1094	2	C59434	KIAA1688 protein [
14	154.5	3.7	1426	2	T30817	homeotic protein C
15	153.5	3.7	2058	2	A59267	myosin X - human
16	152.5	3.7	2052	2	T18519	myosin X - bovine
17	152.5	3.7	2062	2	A59297	myosin X - mouse
18	148.5	3.6	555	2	T30349	structural protein
19	139.5	3.4	1316	2	T00381	KIAA0633 protein -
20	138	3.3	449	2	S16748	proline-rich prote
21	135.5	3.3	534	2	S21961	proline-rich prote
22	135	3.3	1420	2	T37781	probable cytoskele
23	134	3.2	956	2	T00247	zinc finger protei
24	134	3.2	1468	2	S11515	formin - mouse
25	134	3.2	1561	2	T00248	zinc finger protei
26	133.5	3.2	1014	2	T13476	hypothetical prote
27	133	3.2	699	2	T05225	extensin homolog F
28	133	3.2	3938	2	T42761	Bassoon protein -
29	132	3.2	1206	2	S24407	formin isoform 1V
30	131.5	3.2	1466	2	T17138	CL1AA protein - ra
31	131.5	3.2	1471	2	T17149	CL1BA protein - ra
32	131.5	3.2	1510	2	T17145	CL1AB protein - ra
33	131.5	3.2	1515	2	T17156	CL1BB protein - ra
34	130.5	3.2	1032	2	D83637	serine/threonine p
35	130.5	3.2	1201	2	G86441	unknown protein [i
36	130	3.1	1357	2	T29265	hypothetical prote
37	129.5	3.1	3942	2	T42730	Bassoon protein -
38	129	3.1	533	2	S37781	retinoid X recepto
39	128	3.1	727	2	C84534	hypothetical prote
40	128	3.1	1188	2	S49915	extensin-like prot
41	128	3.1	1281	2	T00346	hypothetical prote
42	127.5	3.1	228	2	S53504	extensin-like prot
43	127.5	3.1	263	2	T12698	extensin - common
44	127.5	3.1	1467	2	T18411	latrophilin-1, bra
45	127.5	3.1	1472	2	T18413	latrophilin-1, bra

ALIGNMENTS

RESULT 1
A59295
unconventional myosin-15 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59295
R;Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W., Probst, F.J.; Barber, T.D.; Mil
an, T.B.; Fridell, R.A.
Genomics 61, 243-258, 1999
A;Title: Characterization of the human and mouse unconventional myosin XV genes responsi
A;Reference number: A59266; MUID:20021762; PMID:10552926
A;Accession: A59295
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3511 <LIA>
A;Cross-references: GB:AF141035; NID:g6224684; PIDN:AAF05804.1; PID:g6224685
C;Genetics:
A;Gene: MGI:Myo15
A;Cross-references: MGI:1261811
A;Map position: 11:33.9
C;Superfamily: myosin motor domain homology
F;1209-1871/Domain: myosin motor domain homology <MMO>

Query Match 28.1%; Score 1159.5; DB 2, Length 3511;
Best Local Similarity 34.3%; Pred. No. 4.6e-69;
Matches 289; Conservative 136; Mismatches 245; Indels 173; Gaps 22;

```
Qy       4 SPPGFVPVPVQPSRPPPKAPLPKIDPPTDEALAKLGINGAHSSPPMLSPGPGKGPPPAVAPR 63
           |||      |: |   ||  :                  |  : | :: |:|    | :||
Db    2405 SPPQLPPPPLAPEAPPKTVVG--------------TGPPAKPVLVRPTPQSWAPGSVAKA 2510

Qy      64 PKAP----------------LQLGPS---------- ---     SSIKE-KQGP-  82
           | | |                :   |                      | ||: || |
Db    2511 PKIPSKPVAVPILAQDWTAPESISASPELVRYSTLNSEHPLQPTQQIRSIFQVKQPPWA 2570

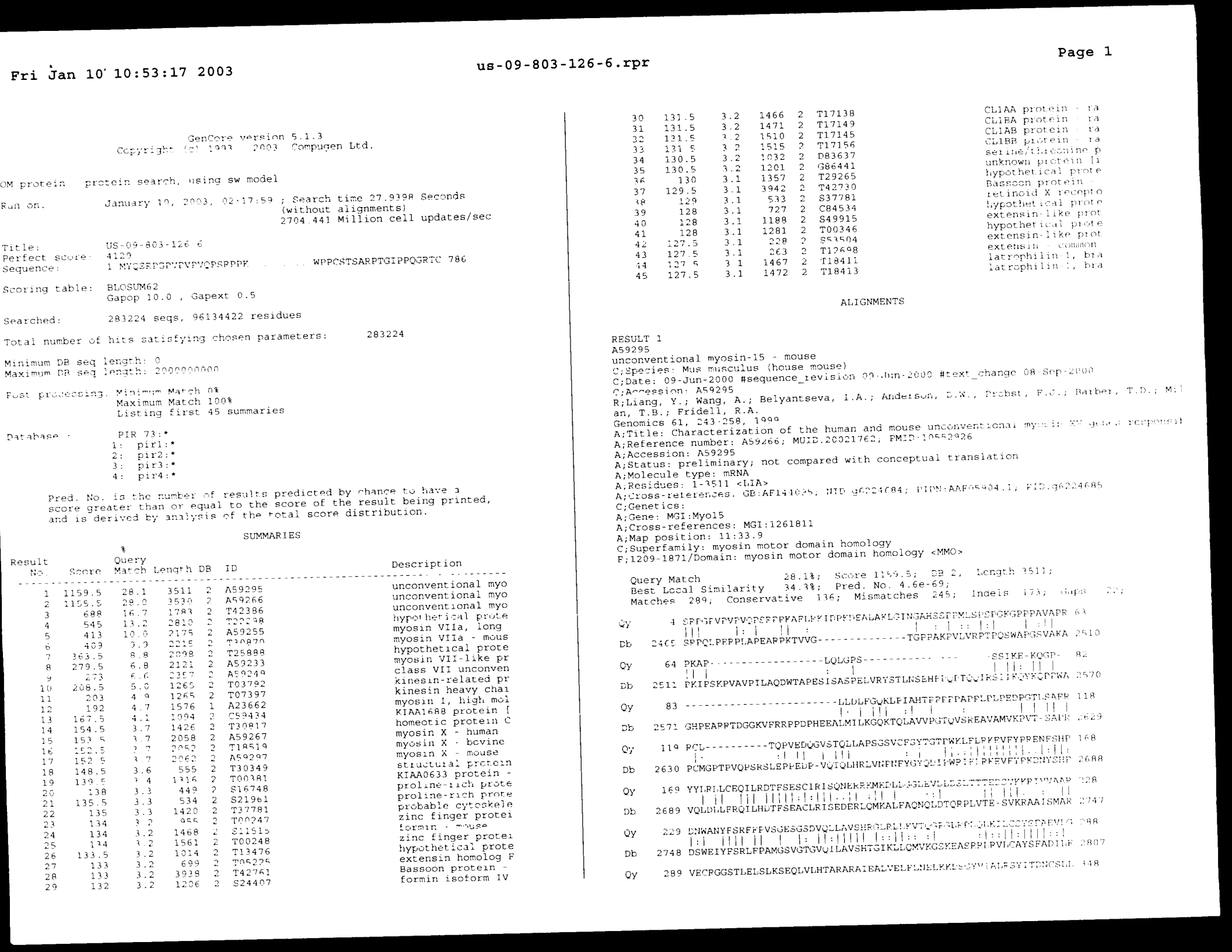
Qy      83 ----------------------LLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAPR 118
                                 |: | |||  :|  |    :              | | | |
Db    2571 GHPEAPPTDGGKVFRRPPDPHEEALMILKGQKTQLAVVPGTQVSREAVAMVKPVT-SAPR 2629

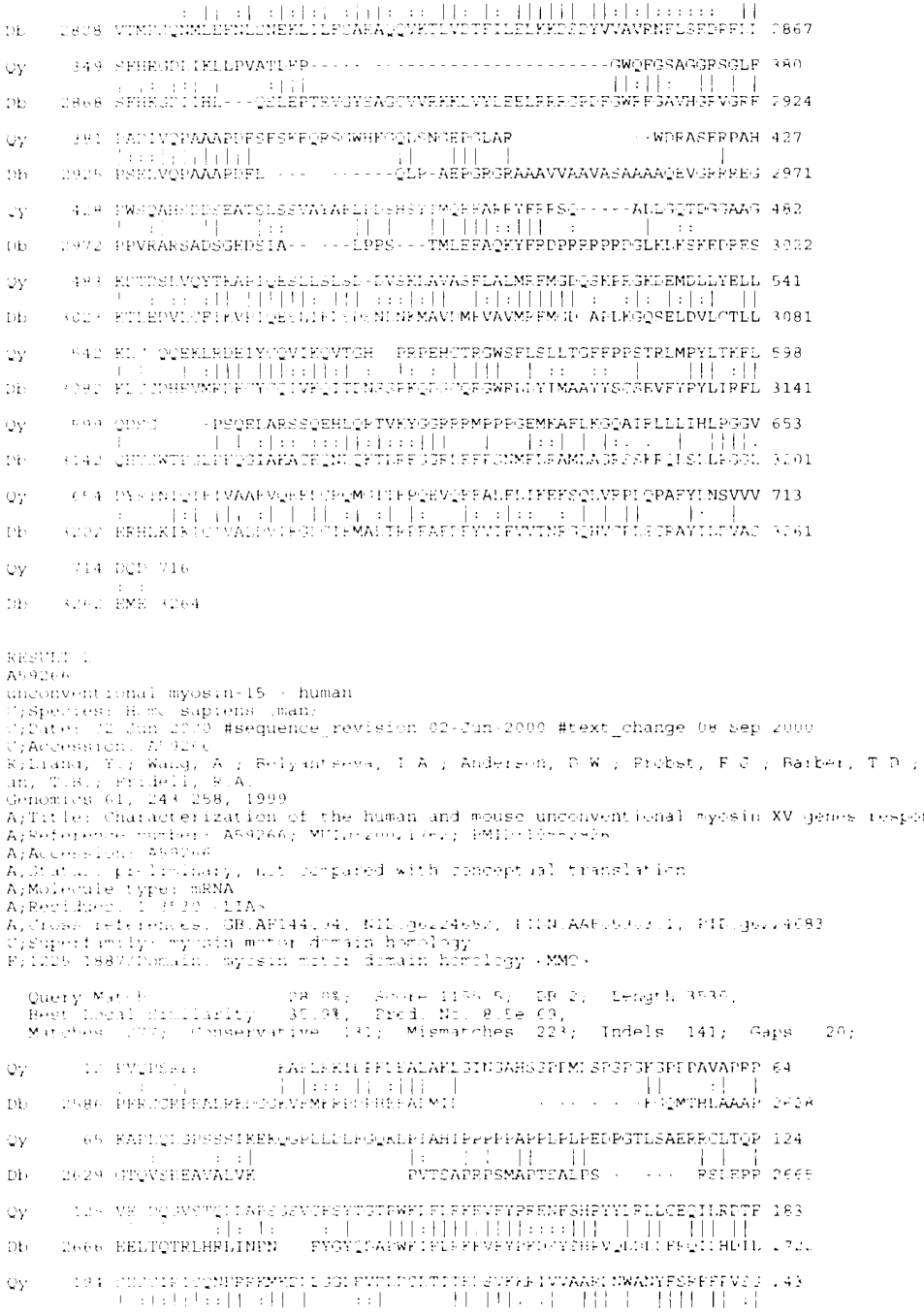
Qy     119 PCL----------TQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLPLPKPVFYPPENFSHP 168
           |.           :| ||  | |||          : |   ||.:|||||||||..|:|||
Db    2630 PCMGPTPVQPSRSLEPPEDP-VQTQLHRLVNFNFYGYQQIPWPIPLPKPVFYPPKDNYSHP 2688

Qy     169 YYLPLLCFQILRDTFSESCIRISQNEKRFMKDLLSGLEVLDLSLTTTEDQVKPPIVWAAP 228
           | ||  ||| |||||:|:|||..|| :|| |    :|     || |||. : ||
Db    2689 VQLDLLFRQILHDTFSEACLRISEDERLQMKALFAQNQLDTQRPLVTE-SVKRAAISMAR 2747

Qy     229 DNWANYFSRFFPVSGESGSDVQLLAVSHRGLRLLFVTQGPNLRPQLLKILCGYSFPEVIG 288
           |:|  |||| ||   |  |: ||:||||| |::||:: :|      :|::||:||||::|
Db    2748 DSWEIYFSRLFPAMGSVGTGVQILAVSHTGIKLLQMVKGSKEASPPIPVLCAYSFADIHF 2807

Qy     289 VECPGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKLDSGVVIALFSYITDNCSLL 348
```





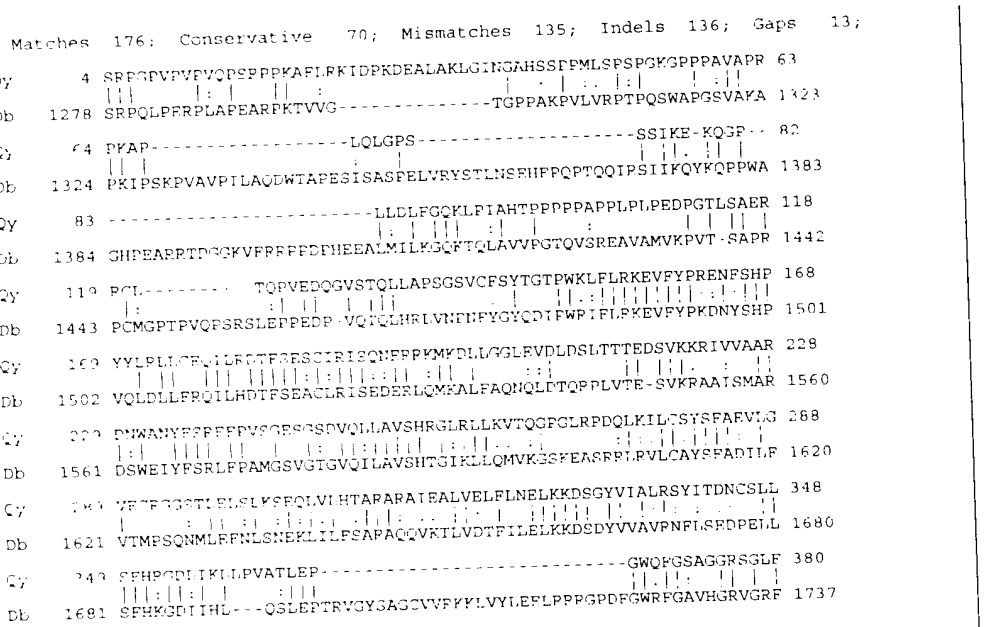
RESULT 1
A59266
Unconventional myosin-15 - human
C;Species: Homo sapiens (man)
C;Date: 12-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59266
R;Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; Mi
an, T.B.; Friedli, R.A.
Genomics 61, 243-258, 1999
A;Title: Characterization of the human and mouse unconventional myosin XV genes respons
A;Reference number: A59266; MUID:20001167; PMID:10534406
A;Accession: A59266
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3529 <LIA>
A;Cross-references: GB:AF144104; NID:g6224682; PIDN:AAF04733.1; PID:g6224683
C;Superfamily: myosin motor domain homology
F;1025-1987/Domain: myosin motor domain homology <MMO>

  Query Match 58.8%; Score 1150.5; DB 2; Length 3529;
  Best Local Similarity 35.0%; Pred. No. 8.5e-69;
  Matches 1077; Conservative 742; Mismatches 223; Indels 141; Gaps 20;

RESULT 2
T42386
unconventional myosin 15 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
C;Accession: T42386; JE0268
R;Probst, F.J.; Fridell, R.A.; Raphael, Y.; Saunders, T.L.; Wang, A.; Liang, Y.; M
Science 280, 1444-1447, 1998
A;Title: Correction of deafness in shaker-2 mice by an unconventional myosin in a BA
A;Reference number: Z22145; MUID:98267310; PMID:9603739
A;Accession: T42386
A;Status: preliminary; translated from GP/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1793 <PPO>
A;Cross-references: EMBL:AF051130; NID:g3108807; PID:g3108808; PIDN:AAC04014.1
R;Wakabayashi, Y.; Takahashi, Y.; Kikkawa, Y.; Okano, H.; Mishima, Y.; Ushiki, T.; Y
Biochem. Biophys. Res. Commun. 248, 655-659, 1998
A;Title: A novel type of myosin encoded by the mouse deafness gene shaker-2.
A;Reference number: JE0268; MUID:98369604; PMID:9703981
A;Accession: JE0268
A;Molecule type: mRNA
A;Residues: 70-162,'KI',163-409,'R',411-654 <WAK>
A;Cross-references: DBRS:AP014518; NID:g4127466; PIDN:BAA33842.1; PID:d1033469; PIDs:
A;Note: this sequence variant may cause mouse deafness
C;Genetics:
A;Gene: Myo15; shaker-2
A;Map position: 11
C;Function:
A;Description: may be required for actin organization in the hair cells of the ou
C;Superfamily: myosin motor domain homology
F;4-702/Domain: myosin motor domain homology <MMO>

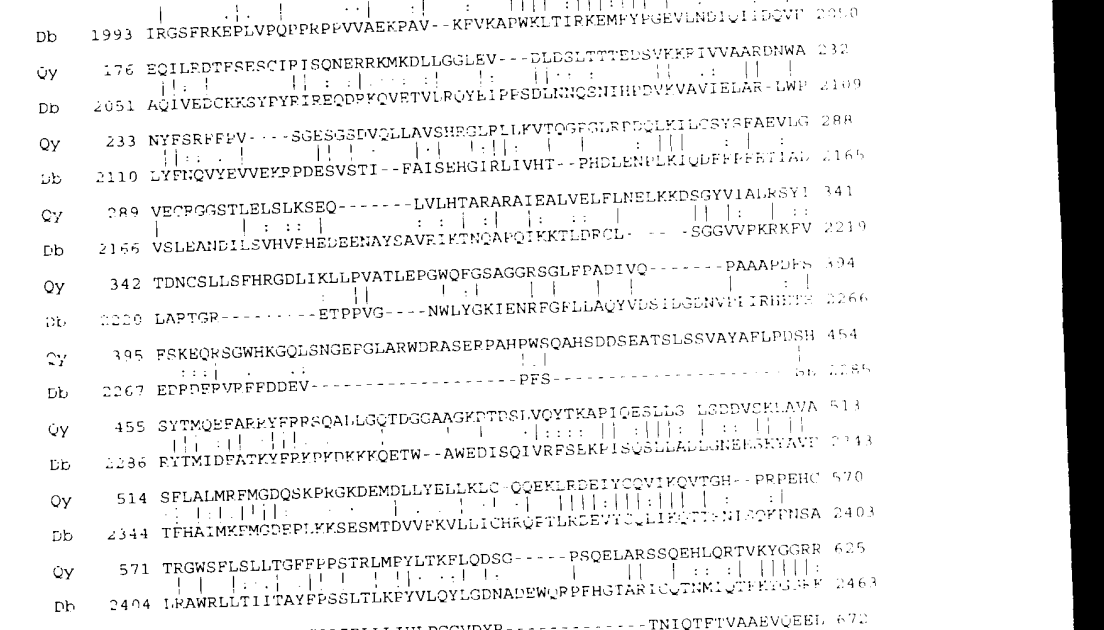
  Query Match 34.0%; Score 668; DB 2; Length 1793;
  Best Local Similarity 34.0%; Pred. No. 9.2e-38;

Matches 176; Conservative 70; Mismatches 135; Indels 136; Gaps 13;



RESULT 4
T22298
hypothetical protein F46C3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18 Feb-2000
C;Accession: T22298
R;Cottage, A.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19543
A;Accession: T22298
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2810 <WIL>
A;Cross-references: EMBL:Z66563; PIDN:CAA91463.1; GSPDB:GN00028; CESP:F46C3.3
A;Experimental source: clone F46C3
C;Genetics:
A;Gene: CESP:F46C3.3
A;Map position: X
A;Introns: 67/2; 172/3; 294/2; 310/3; 366/2; 431/2; 684/3; 739/3; 786/3; 803/2; 880/2; 1
A;Introns: 1958/3; 1999/1; 2078/1; 2113/3; 2159/2; 2220/1; 2269/3; 2306/1; 2399/2; 2444/1; 2488/3;

Query Match 13.2%; Score 545; DB 2; Length 2810;
Best Local Similarity 24.2%; Pred. No. 7.4e-28;
Matches 185; Conservative 127; Mismatches 294; Indels 158; Gaps 24;



RESULT 5
A59255
myosin VIIa, long form - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02 Feb 2001
C;Accession: A59255; S71932; S60263
R;Chen, Z.Y.; Hasson, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M.
Genomics 36, 440-448, 1996
A;Title: Molecular cloning and domain structure of human myosin-VIIa, the gene product de
A;Reference number: A59255; MUID:97038686; PMID:8884267
A;Accession: A59255
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-2175 <CHE>
A;Cross-references: GB:U55208; NID:g1613787; PIDN:AAC50427.1; PID:g1613788
A;Experimental source: tissue type testis
R;Weil, D.
submitted to the EMBL Data Library, November 1994
A;Reference number: S71932
A;Accession: S71932
A;Molecule type: mRNA
A;Residues: 96-334,'T',336-370,'N',372-388,'VTG',392-469,'F',471-563 <WEI>
A;Cross-references: EMBL:U17180; NID:g1235669
A;Note: the sequence is revised in GenBank entry HSU39226, release 114, (PIDN:AAB03670.1)
R;Weil, D.; Blanchard, S.; Kaplan, J.; Guilford, P.; Gibson, F.; Walsh, J.; Mburu, P.; Va
Piet, D.; Munnich, A.; Steel, K.P.; Brown, S.D.M.; Petit, C.
Nature 374, 60-61, 1995
A;Title: Defective myosin VIIA gene responsible for Usher syndrome type 1B.
A;Reference number: S60263; MUID:95174880; PMID:7870171

A;Accession: S60264
A;Molecule type: mRNA
A;Residues: 96-157;199-255 <WEW>
A;Cross-references: EMBL:U17198; NID:g1735[illegible]
A;Note: the sequence is revised in GenBank entry U17198, release 114, (FIRST AAB03679.1)
C;Comment: A alternate transcript from the same gene is shown in PIR:A53257.
C;Genetics:
A;Gene: MYO7A
A;Map position: 11q13.5
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; ATP; nucleotide binding; P-loop
F;96-865/Domain: myosin motor domain homology <MMOT>
F;158-165/Region: nucleotide-binding motif A (P-loop)

Query Match 10.0%; Score 413; DB 2; Length 2175;
Best Local Similarity 24.0%; Pred. No. 3.9e-19;
Matches 197, Conservative 112, Mismatches 265, Indels 216, Gaps 30,

Qy 39 DEALAELGINGAHSSPPM-- LSPSPGKGP PPAVAPPPKAPQLGPGSSSIPE 98
Db 1228 EERLRPTPVNGTRTQPPGWLELQATKCPPPIMLPVTPMCSTTPTTLCTCAATTAPPLCNAL 1287

Qy 99 -PQGPLLLLPGLPLPIAHTPPPPPAPPLPLPPPPGTLSAPPPCTTQPVPDQPVSTQLLAP 157
Db 1288 ADKISIKPPPGPSLYIA---------- --- - - - -PPPVPGSS bigpVPMD 1341

Qy 158 SGSVCPGYT --- -GTPWKLPLRKPPVPTPPPENPGHP TTLRLLYEPLILAVTPPENT 208
Db 1322 AISQCEQYAPEQGAQEPNAPWPLPPPKEVPTPWHNPGEDNVATNLIYQQVVPG V 1377

Qy 209 PISQNPPPKMKDLLG--------- GLPVDLDSL - - 252
Db 1376 PPGEYPCPPPPNGLAELASQQYPVDYGSPMIIPRPISIVPIYPGPPPIGANLPPWAQLA 1435

Qy 213 ---------- TTTPDSVKKRIVVAAPDNWANYPPPPPVPSPPP PGVQLLAVS 255
Db 1436 IAAHKPGIYAQPRPIWQPVKLDVVNYAPPPWMPIPSKEYPAYPP- -PPIPP V DVAVN 1494

Qy 256 HRGLPLLPVTQGPGLPPGQLPILCTYSPAEVLGV - EGR GGSTLE 298
Db 1495 WTGVYPV- - - - DEQEQVLLELGPTPEIMAVGSSKEGRVILSLGPSDLGPATPQSGR 1549

Qy 299 ---------- LSLKSFQLVIHTAKARATPALVELPLNELKKIS 331
Db 1546 AGLTPAGPCSPCWSCRGTPMMAPSPTLATTRGREYTPSSNAELPS..VVTPLEGLPRPG 1600

Qy 332 GYVIALPSYITDN---- --CSLLSPHRGDLIKLLPVATLEPGWLPPGAAG PPS 378
Db 1606 KYVVALQ----DNPNPACPPSGPLSFAPGDLI-ILDHDTGEQVMNSGWANGINERTKQPG 1660

Qy 379 LPFAD IVQPAA---------- - APDPSPGREQRSGWHPSQLDNGIEPSLARWDPAAPE 431
Db 1661 LPFGG IVQPAA---------- - APDPSPGREQRSGWHPSQLDNGIEPSLARWDPAAPE 1715

RESULT 6
T30870

myosin VIIa - mouse
N;Alternate names: gene shaker-1 protein
C;Species: Mus musculus (house mouse)
I;Date: 01-Oct-1999 #sequence_revision 07-[illegible]-1999 #text_change 02-Feb-[illegible]
I;Accession: T30870; S60346
R;Walsh, J., Mburu, P., Gibson, F., Liu, X., Law, I., Steel, K.P., Brown, S.D.M.
submitted to the EMBL Data Library, December 1996
A;Description: Mutation analysis of the mouse myosin VIIa deafness gene.
A;Reference number: Z20913
A;Accession: T20870
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2215 <WAL>
A;Cross-references: EMBL:U81453; NID:g1778381; PID:g1778382; PIDN:AAB49708.1
R;Gibson, F.; Walsh, J.; Mburu, P.; Varela, A.; Brown, K.A.; Antonio, M.; Beisel, K.
Nature 374, 62-64, 1995
A;Title: A type VII myosin encoded by the mouse deafness gene shaker-1
A;Reference number: S60346; MUID:95174881; PMID:7870171
A;Accession: S60346
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 96-156,'R',158-237,'E',239-681,'V',683-716 <GIB>
A;Note: this sequence variant is from the mouse deafness gene shaker-1
C;Genetics:
A;Gene: sh1; MyoVIIa; MGI:MyoVIIa
A;Cross-references: MGI:104501
A;Map position: 7 - 48.1
C;Function:
A;Description: seems to be involved in auditory transduction
C;Superfamily: myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;68-720/Domain: myosin motor domain homology <MMOT>
F;96-716/Domain: myosin motor domain homology (fragment) <MMOT>
F;158-165/Region: nucleotide binding motif A (P-loop)

Query Match 9.9%; Score 409; DB 2; Length 2215;
Best Local Similarity 23.2%; Pred. No. 7.4e-19;
Matches 190, Conservative 113, Mismatches 261; Indels 204; Gaps 21;

Qy 39 DEALAELGINGAHSSPPM-- LSPSPGKGP PPAVAPPPKAPQLGPGSSSIPE 98
Db 1228 EERLRPTPVNGTRTQPPGWLELQATKCPPPIMLPVTPMCSTTPTTLCTCAATTAPPLCNAL 1287

Qy 99 -PQGPLLLLPGLPLPIAHTPPPPPAPPLPLPPPPGTLSAPPPCTTQPVPDQPVSTQLLAP 157
Db 1288 ADKISIKPPPGPSLYIA---------- --- - - - -PPPVPGSS bigpVPMD 1341

Qy 158 SGSVCPGYT --- -GTPWKLPLRKPPVPTPPPENPGHP TTLRLLYEPLILAVTPPENT 208
Db 1322 AISQCEQYAPEQGAQEPNAPWPLPPPKEVPTPWHNPGEDNVATNLIYQQVVPG V 1377

Qy 209 PISQNPPPKMKDLLG--------- GLPVDLDSL - - 252
Db 1376 PPGEYPCPPPPNGLAELASQQYPVDYGSPMIIPRPISIVPIYPGPPPIGANLPPWAQLA 1435

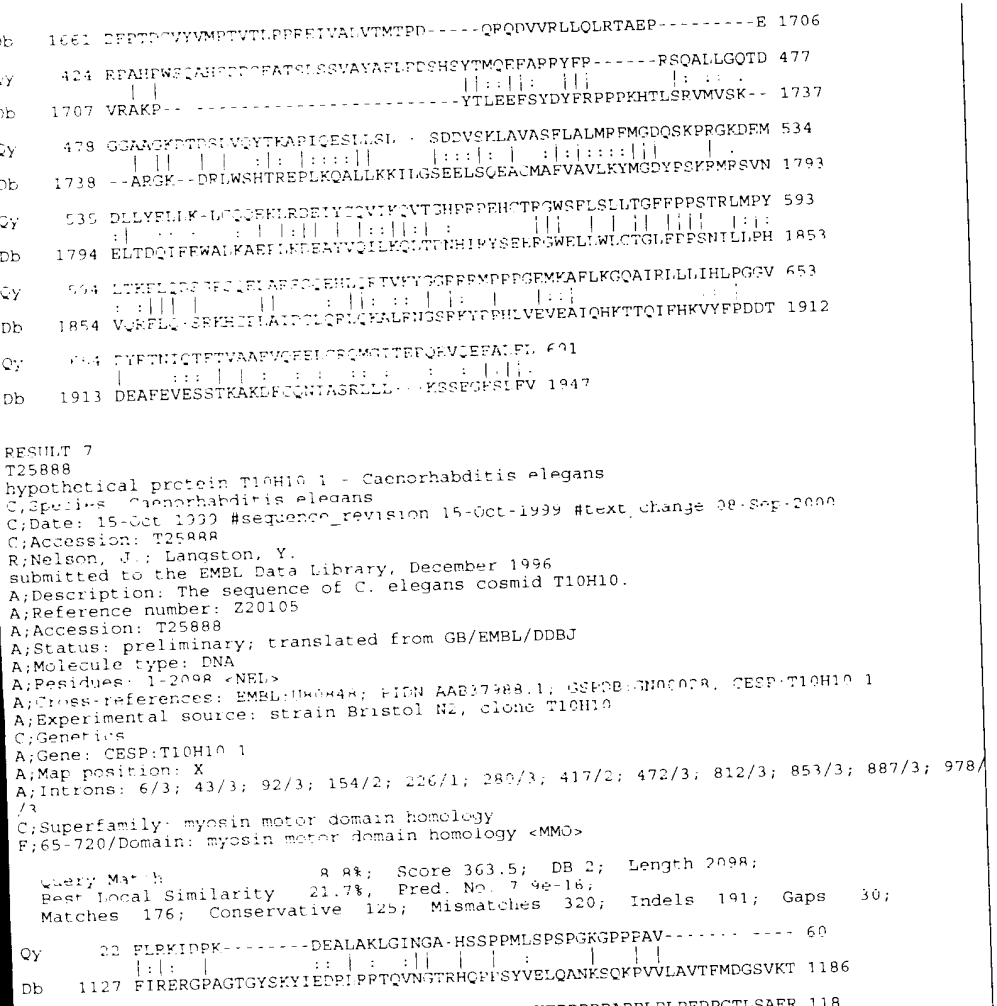
Qy 213 ---------- TTTPDSVKKRIVVAAPDNWANYPPPPPVPSPPP PGVQLLAVS 255
Db 1436 IAAHKPGIYAQPRPIWQPVKLDVVNYAPPPWMPIPSKEYPAYPP- -PPIPP V DVAVN 1494

Qy 256 HRGLPLLPVTQGPGLPPGQLPILCTYSPAEVLGV - EGR GGSTLE 298
Db 1495 WTGVYPV- - - - DEQEQVLLELGPTPEIMAVGSSKEGRVILSLGPSDLGPATPQSGR 1549

Qy 299 ---------- LSLKSFQLVIHTAKARATPALVELPLNELKKIS 331
Db 1546 AGLTPAGPCSPCWSCRGTPMMAPSPTLATTRGREYTPSSNAELPS..VVTPLEGLPRPG 1600

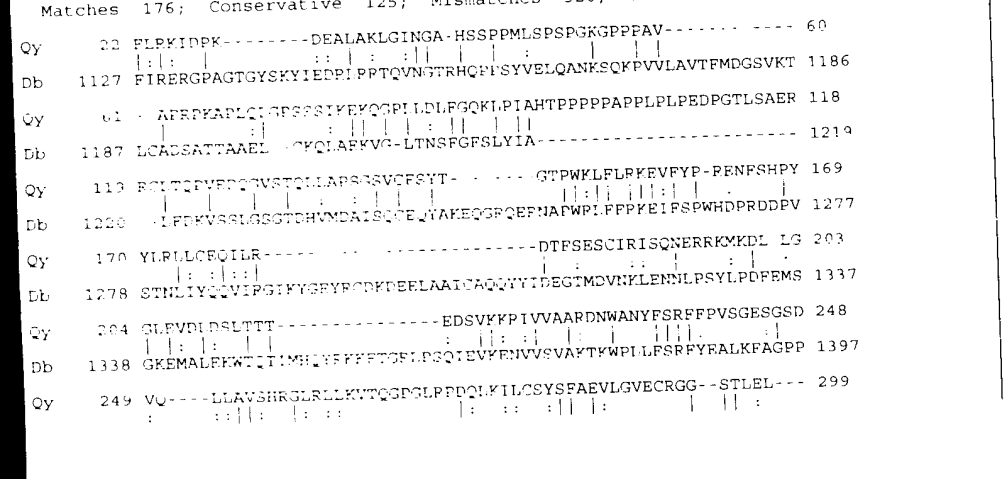
Qy 332 GYVIALPSYITDN---- --CSLLSPHRGDLIKLLPVATLEPGWLPPGAAG PPS 378
Db 1606 KYVVALQ----DNPNPACPPSGPLSFAPGDLI-ILDHDTGEQVMNSGWANGINERTKQPG 1660

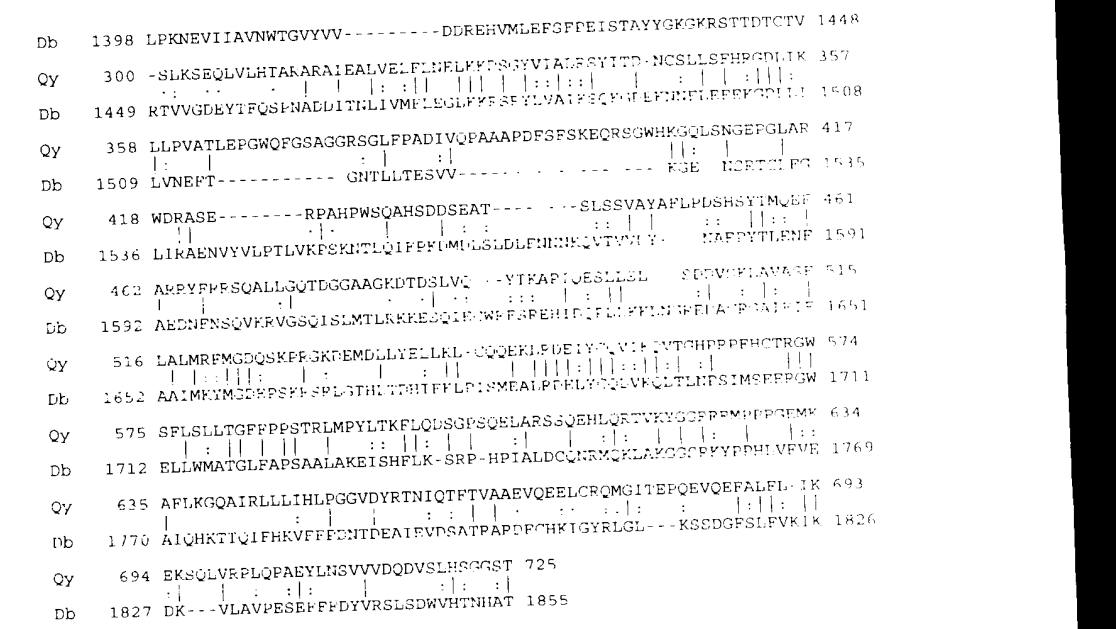
Qy 379 LPFAD IVQPAA---------- - APDPSPGREQRSGWHPSQLDNGIEPSLARWDPAAPE 431



RESULT 7
T25888
hypothetical protein T10H10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T25888
R;Nelson, J.; Langston, Y.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid T10H10.
A;Reference number: Z20105
A;Accession: T25888
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2098 <NEL>
A;Cross-references: EMBL:U80848; PIDN:AAB37488.1; GSPDB:GN00028; CESP:T10H10.1
A;Experimental source: strain Bristol N2, clone T10H10
C;Genetics
A;Gene: CESP:T10H10.1
A;Map position: X
A;Introns: 6/3; 43/3; 92/3; 154/2; 226/1; 280/3; 417/2; 472/3; 812/3; 853/3; 887/3; 978/3
C;Superfamily: myosin motor domain homology
F;65-720/Domain: myosin motor domain homology <MMO>

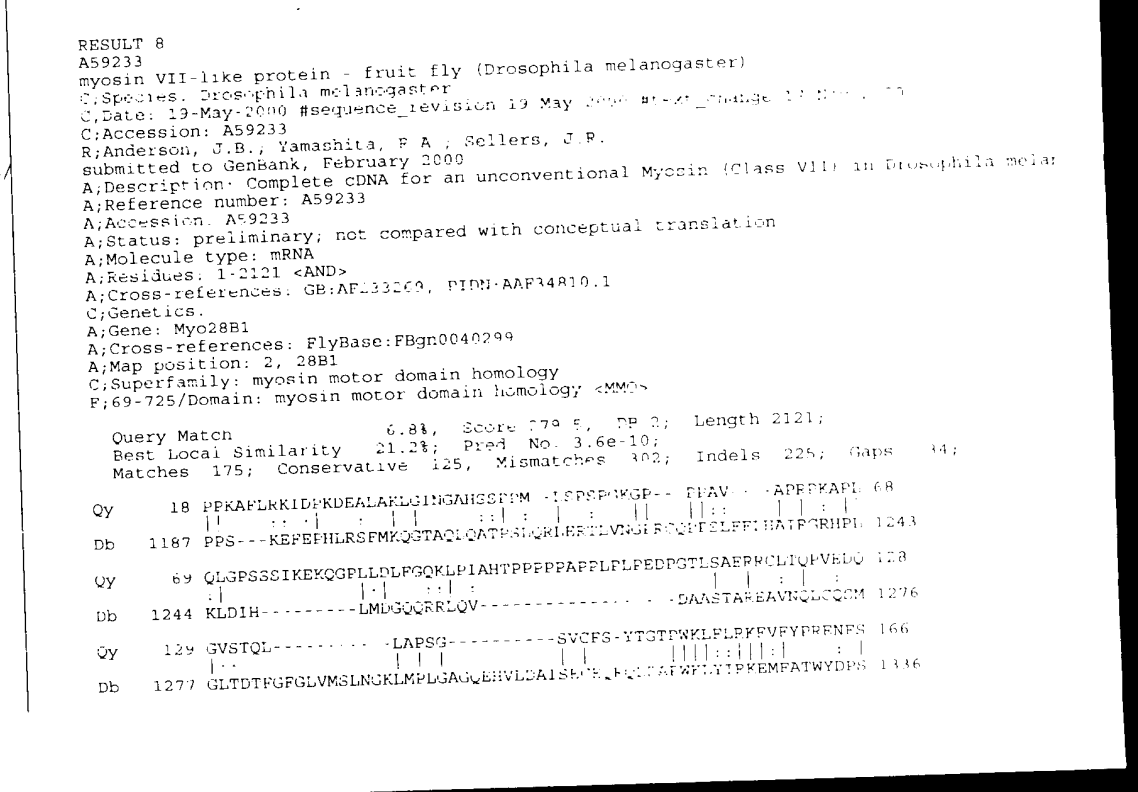
Query Match 8.8%; Score 363.5; DB 2; Length 2098;
Best Local Similarity 21.7%; Pred. No. 7.9e-16;
Matches 176; Conservative 125; Mismatches 320; Indels 191; Gaps 30;

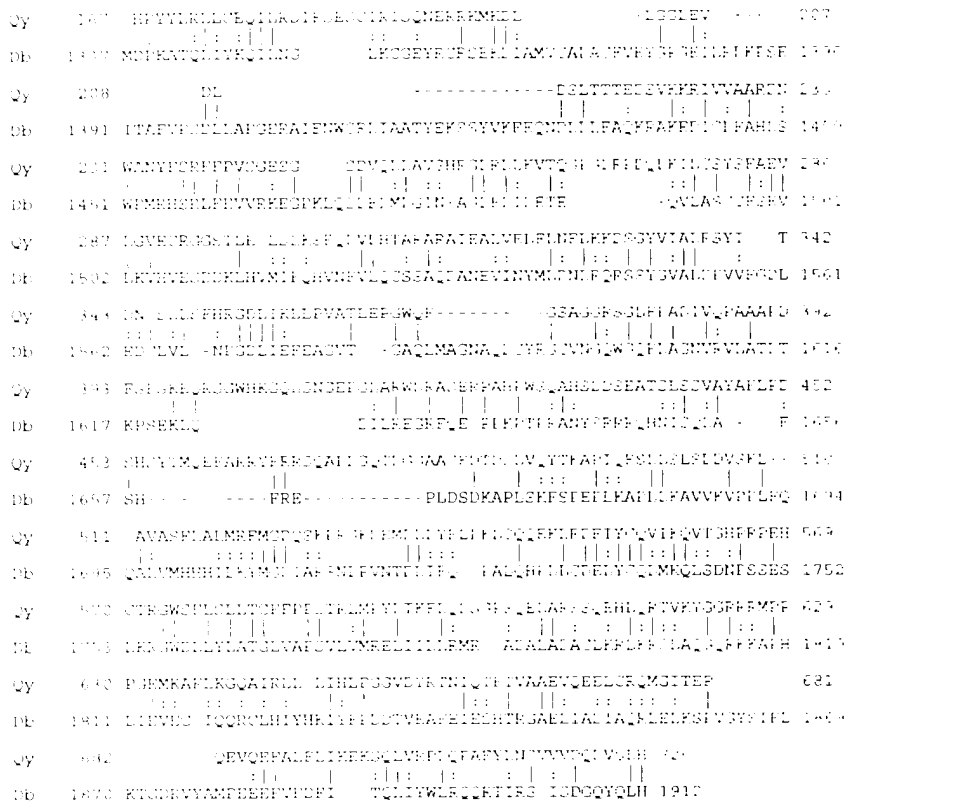




RESULT 8
A59233
myosin VII-like protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: A59233
R;Anderson, J.B.; Yamashita, F.A.; Sellers, J.R.
submitted to GenBank, February 2000
A;Description: Complete cDNA for an unconventional Myosin (Class VII) in Drosophila melanogaster
A;Reference number: A59233
A;Accession: A59233
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-2121 <AND>
A;Cross-references: GB:AF133209; PIDN:AAF34810.1
C;Genetics.
A;Gene: Myo28B1
A;Cross-references: FlyBase:FBgn0040299
A;Map position: 2, 28B1
C;Superfamily: myosin motor domain homology
F;69-725/Domain: myosin motor domain homology <MMO>

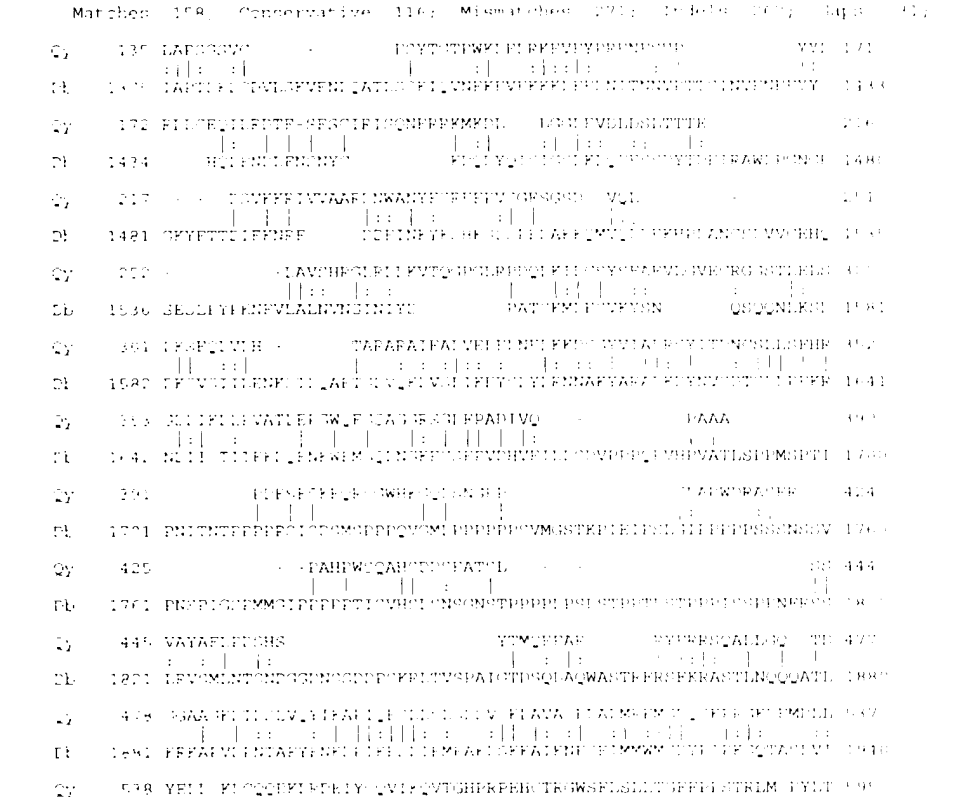
Query Match 6.8%; Score 279.5; DB 2; Length 2121;
Best Local Similarity 21.2%; Pred. No. 3.6e-10;
Matches 175; Conservative 125; Mismatches 302; Indels 225; Gaps 34;





RESULT 9
A59249
class VII unconventional myosin - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
C;Accession: A59249
R;Titus, M.A.; Kuspa, A.; Loomis, W.F.
Proc. Natl. Acad. Sci. U.S.A. 91, 9436-9440, 1994
A;Title: Discovery of myosin genes by physical mapping in Dictyostelium.
A;Reference number: Z20873; MUID:95023928; PMID:7937787
A;Accession: A59249
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-2457 <TIT>
A;Cross-references: GB:L35321; NID:g6226760; PIDN:AAF06935.1; PID:g6226761
A;Experimental source: gene myo1; product class VII unconventional myosin
R;Titus, M.A.
Curr. Biol. 9, 1297-1303, 1999
A;Title: A class VII unconventional myosin is required for phagocytosis
A;Reference number: A59249
A;Contents: annotation
C;Genetics:
A;Gene: myo1
C;Superfamily: myosin motor domain homology <MMO>
F;16-676/Domain: myosin motor domain homology <MMO>

Query Match 6.6%; Score 273; DB 2; Length 2357;
Best Local Similarity 19.6%; Pred. No. 1.1e-09;



RESULT 10
T03792
kinesin-related protein tck1 - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2001
C;Accession: T03792
R;Wang, W.; Takezawa, D.; Narasimhulu, S.B.; Reddy, A.S.N.; Poovaiah, B.W.
Plant Mol. Biol. 31, 87-100, 1996
A;Title: A novel kinesin-like protein with a calmodulin binding domain.
A;Reference number: Z15089; MUID:96309387; PMID:8704162
A;Accession: T03792
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1265 <WAN>

```
A;Cross-references: EMBL:U52078; NID:g1491930; PIDN:AAC49393.1; PID:g1491931
A;Experimental source: cultivar Petit Havana SR-1; tissue-type anthers
C;Genetics:
A;Gene: tck1
C;Superfamily: kinesin motor domain homology
F;890-1221/Domain: kinesin motor domain homology <KMOT>

  Query Match                5.6%;  Score 209.5;  DB 2;  Length 1265;
  Best Local Similarity   23.3%;  Pred. No. 1e-05;
  Matches   77;  Conservative   66;  Mismatches  130;  Indels   57;  Gaps   10;

Qy     409 SNGEPGLARWDRAS-------ERPAH-PWSQAHSDDSEA----------TSLSSVAYAFL 450
           |:  |    |  |:|       |||:|  :: :: || ::           |:||||    |
Db       3 SDMPPVSMRSSRSSFGSSNGYERPSHYSFATSNGDDYDSDGSNFAPPTPTTLSSVLSPEL 62

Qy     451 PDSHSYTMQEFARFYFRRSQALLSQTPGGAAGK  ... ... -------DTDSLVQYTK 493
           :  |  :      : :  | |     :|||                        : :: : :
Db      63 AGAIPYIDRFQVEGFLKAMQKQL- - -QSAGKPGFFLKKSVGPQVRFKFTFEDMLCFQR 117

Qy     494 APIQESLLSLSDDVSKLAVASFLALMPFMQ   -DQSKFPGKDE MDLLYELLK-LCQQE 547
           ||  |:| :: |:     |  |    ..|    |:: |   || ::|: :| |   ::
Db     118 EPIPTSILKINGDLVGRTVKLFQNILKYMGIDSYDRAAPISLDERIELVGKLFKQALKPS 177

Qy     548 KLPDEIYCQVIKQVTGHPRPEHCTRGWSFLSLLTGFFPPSTRLMPYLTKF-------LQD 600
           :|||||  |: ||   :|     :  | :|     ||   . ||:::          :
Db     178 ELPDEMFAQISKQTRNNPERHSLIKAWFLMYLCASCMPPSKEIGGYLSEYIHTVAHGINT 237

Qy     601 SGPSQELARSSQEHLQRTVEYCGGFFMPPDGPMKAFLKGQAIRLLLIHLPGGVDYRTNIQ 660
           |  | ::   |:|:|:| | |  :|    |::| | |: :  ::  |     :  |
Db     238 TAEVQVVAINTLNALKRSIKAGPPHTIPYEFEIEALLTGKKLTTIVFFLDETFERITYDM 297

Qy     661 TFTVAAEVQEELCPQMGITPEQEVQEFALF 690
           |||  ::|     || :     |:||
Db     298 ATTVADAIEE - -VAGIIKLSAHASFSLF 323


RESULT 11
T07397
kinesin heavy chain-like protein (clone PKCBP) - potato
C;Species: Solanum tuberosum (potato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 12 May 2000
C;Accession: T07397
R;Reddy, A.S.; Narasimhulu, S.B.; Safadi, F.; Golovkin, M.
Plant J. 10, 9-21, 1996
A;Title: A plant kinesin heavy chain-like protein is a calmodulin-binding protein.
A;Reference number: Z16008; MUID:96314858; PMID:8758976
A;Accession: T07397
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1265 <RED>
A;Cross-references: EMBL:L46702; NID:g1369851; PIDN:AAB37756.1; PID:g1369852
C;Superfamily: kinesin motor domain homology
C;Keywords: GTP binding; microtubule binding
F;890-1216/Domain: kinesin motor domain homology <KMOT>

  Query Match                4.9%;  Score 203;  DB 2;  Length 1265;
  Best Local Similarity   24.7%;  Pred. No. 2.4e-05;
  Matches   75;  Conservative   56;  Mismatches  133;  Indels   46;  Gaps    9;

Qy     423 ERPAH---PWSQAHSDDSEAT-----  - SLSSVAYAFLPDSHSY----------TMQE 460
           | |:|    | |          ||: :         :||||    |   :  |       ||:
Db      24 ETPSHYSFPTSNGDDYDSDGSNFAPPTPNTLSSVLSPELAGAIPYIDRFQVEGFLKAMQK 83

Qy     461 FARRYFRPSQALLSQTPGGAAGKDT-DSLVQYTKAPIQESLLSLSDDVSKLAVASFLALM 519
           :|  |          | | : :: : : ||  |:| |: |:   ||  | :::
Db      84 QIHSASHSGFFLKKSVGPQVRPKFTFEDMLCFQRPIPTSILKLNGDLISRAVKLFQSIL 143

Qy     520 KFMS - - DQSKFPGKDE MDLLYELLK-LCQQEKLPDEIYCQVIKQVTGHPRPEHCTRG 573
           ::||    |:  |   || ::|:  | |    :  :|||:  |  ||    :
Db     144 KYMGIDSYDRVAFISLDERIELVGKTFKQALKPSELPDEMFAQISKQTRNNPERHSLIKA 203

Qy     574 WSFLSLLTGFFPPSTRLMPYLTKFL-------QDSGPSQELARSSQEHLQRTVEYCGGFPP 626
           |  : |     |||  :  ||::::          |  | ::   |:|::| | |
Db     204 WELMYLCASCMPPSKEIGGYLSEYIHTVAHGTNTDSEVQVFAINTLNALKRSIKAGPPHT 263

Qy     627 MPPPGEMKAFLKGQAIRLLLIHLPGGVDYRTNIQTFTVAAEVQEELCPQMGITPEQEVQE 686
           :|   |..| | |: .  .. |     :  |     |||  ::|     ||.
Db     264 IPGREEIEAHLTGKKLTTIVFFLDETFEEITYDMATTVADAIEE----VAGIIKLSAHVS 319

Qy     687 FALF 690
           |:||
Db     320 FSLF 323


RESULT 12
A23662
myosin I, high molecular weight - Acanthamoeba sp.
N;Alternate names: myosin, HMWMI
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Acanthamoeba sp
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19 Apr 2002
C;Accession: A23662
R;Horowitz, J.A.; Hammer III, J.A.
J. Biol. Chem. 265, 20646-20652, 1990
A;Title: A new Acanthamoeba myosin heavy chain. Cloning of the gene and immunological id
A;Reference number: A23662; MUID:91056121; PMID:2243110
A;Accession: A23662
A;Molecule type: DNA
A;Residues: 1-1576 <HGR>
A;Cross-references: GB:J05678
A;Note: the authors translated the codon GAT for residue 280 as Ser, and failed to transl
A;Note: the nucleotide sequence in GB:J05678 differs from that shown in the paper in have
residues 725 and 726
C;Genetics:
A;Introns: 54/3; 105/3; 162/3; 210/1; 324/2; 416/1; 507/1; 644/2; 731/1; 1038/1; 1090/1;
C;Superfamily: myosin I, high molecular weight; myosin motor domain homology, SH3 homolog
C;Keywords: actin binding; ATP; hydrolase; nucleotide binding, P loop
F;78-738/Domain: myosin motor domain homology <MMOT>
F;168-175/Region: nucleotide-binding motif A (P-loop)
F;629-650/Region: actin binding #status predicted
F;783-1575/Domain: carboxyl-terminal <CTD>
F;1445-1510/Region: glutamine-rich
F;1525-1573/Domain: SH3 homology <SH3>
F;174/Binding site: ATP (Lys) #status predicted

  Query Match                4.7%;  Score 192;  DB 1;  Length 1576;
  Best Local Similarity   29.3%;  Pred. No. 0.00018;
  Matches   61;  Conservative   30;  Mismatches   87;  Indels   30;  Gaps    7;

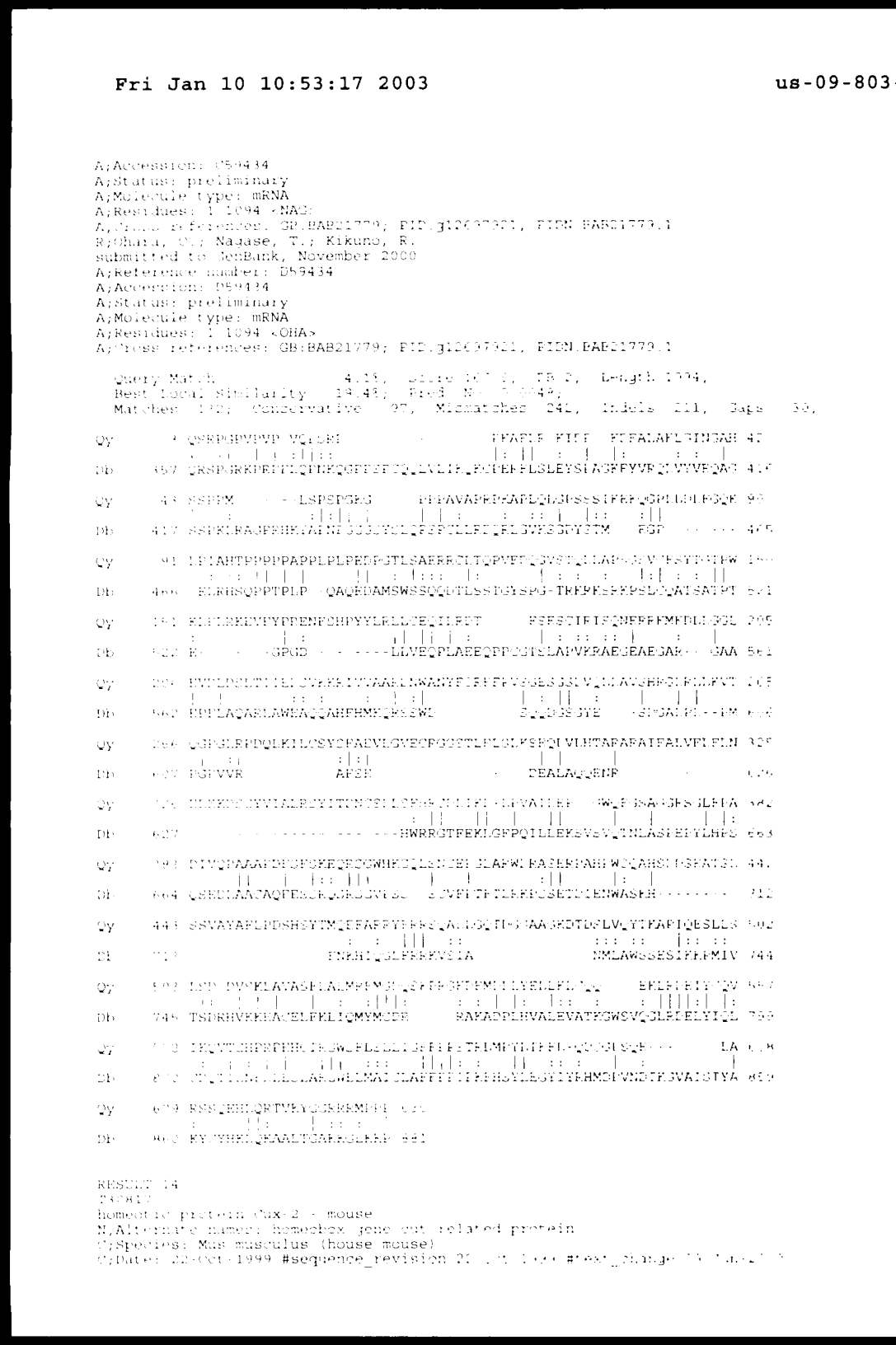
Qy     459 QEFAPPYFPRSQAL-LGQTDGGAAGKDTDSLVQYTKAPIQESLLSLSDDVSKLAVASFLA 517
           :.|:.      :|| ||                 .:||:||  || :|     | |   |
Db    1240 EDFAKAKLLGREALKLG---- - - - --MFKWTKSPIPTSLITLDQDLAVKSATRLPKN 1286

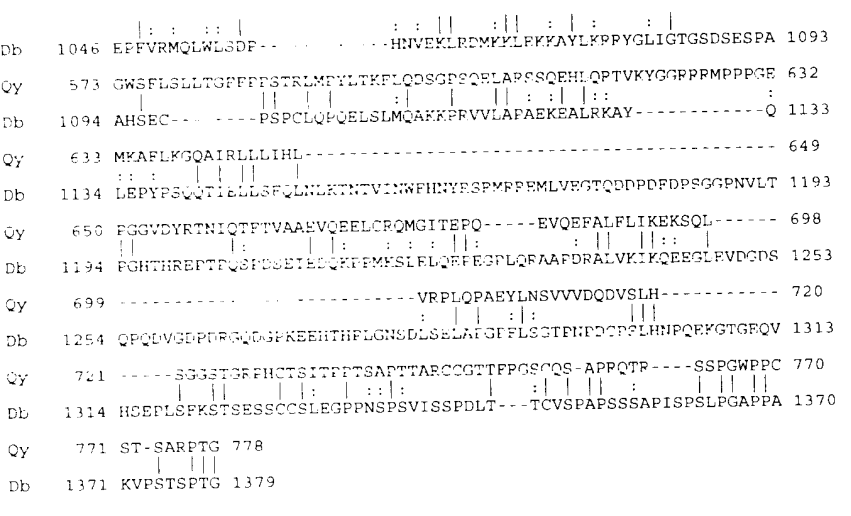
Qy     518 LMRFMGDQSKFPGKDEMDLLYELLKLC QQEKLPDEIYCQVIKQVTGHPRPEHCTRGWSF 576
           .: ||||:  |        |  :||: |    :||:|:|||:|||:| :| |:  |:||
Db    1287 VLGFMGDRPLPYPN---ALAQDLLEQCLAAPELRNEVYCQIIKQTTENFSPQSVTKGWQL 1343

Qy     577 LSLLTGFFPPSTRLMPYLTKFLQDSGPSQELARSSQEHLQRTVEYCGGFPMPPDGPMKAF 636
           :          ||||      |  ||      |  |  | :|| :|   |  |
Db    1344 MRCCLQTFPPSEEFANCLEMFLAAKGKDDKYI----EMLHDT-QYGDKRTSAFNVEQILA 1398

Qy     637 LKGQAIRLLLIHLPGGVDYPTNIQTFTV 664
           |  |:        ::  | . | |
Db    1399 AKQYVTRI----- -DLNQSTQVDTSVV 1419

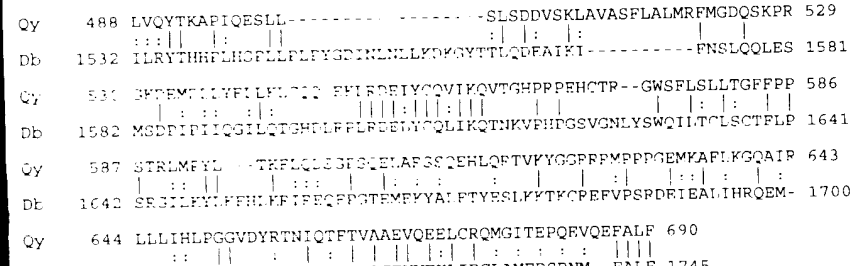

RESULT 13
C59434
KIAA1688 protein [imported] - human
C;Species: Homo sapiens (man)
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 19 Jul 2002
C;Accession: C59434; D59434
R;Nagase, T.; Kikuno, R.; Hattori, A., Kondo, Y., Okumura, K.; Ohara, O.
DNA Res. 7, 347-355, 2000
A;Title: Prediction of the coding sequences of unidentified human genes. XIX. The complet
A;Reference number: C59434
```


A;Accession: D59434
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1094 <NAG>
A;Cross-references: GB:BAB21779; PID:g12697021; PIDN:BAB21779.1
R;Ohara, O.; Nagase, T.; Kikuno, R.
submitted to GenBank, November 2000
A;Reference number: D59434
A;Accession: D59434
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1094 <OHA>
A;Cross-references: GB:BAB21779; PID:g12697021; PIDN:BAB21779.1
Query Match 4.2%, Score 154.5; DB 2, Length 1094;
Best Local Similarity 19.4%; Pred. No. 7.9e-49;
Matches 120; Conservative 97; Mismatches 241; Indels 211; Gaps 30;
RESULT 14
Q78817
homeobox protein Cux-2 - mouse
N;Alternate names: Homeobox jeme cut related protein
C;Species: Mus musculus (house mouse)
C;Date: 01-Oct-1999 #sequence_revision 01-Oct-1999 #text_change 16-Jul-2001
C;Accession: T30817
R;Quaggin, S.E.; Vanden Heuvel, G.B.; Golden, K.; Bodmer, R.; Saneshi, P.
J. Biol. Chem. 271, 22624-22634, 1996
A;Title: Primary structure, neural specific expression, and chromosomal localizat
A;Reference number: Z22582; MUID:96394370; PMID:8798414
A;Accession: T30817
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1426 <QUA>
A;Cross-references: EMBL:U49665; NID:g1518831; PID:g1518832; PIDN:AAC52703.1
C;Genetics:
A;Gene: Cux-2
A;Map position: 5
A;Note: expressed exclusively in the central and peripheral nervous system
C;Superfamily: homeobox protein CTP; cut repeat homology; homeobox homology
C;Keywords: DNA binding, homeobox, nucleus; transcription regulation
F;461-564/Domain: cut repeat homology <CUT1>
F;837-909/Domain: cut repeat homology <CUT2>
F;939-1264/Domain: cut repeat homology <CUT3>
Query Match 3.7%, Score 154.5; DB 2; Length 1426;
Best Local Similarity 18.6%; Pred. No. 0.0054;
Matches 303; Conservative 119; Mismatches 412; Indels 435; Gaps 45;



RESULT 15
A59267
myosin X - human
C;Species: Homo sapiens (man)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59267
R;Rogers, M.S.; Strehler, E.E.
submitted to GenBank, March 2000
A;Description: Identification of myosin X as a specific binding partner for the tumor se
A;Reference number: A59267
A;Accession: A59267
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1 2058 <ROG>
A;Cross-references: GB:AF234532; PIDN:AAF37875.1
C;Genetics:
A;Gene: MYO10
C;Superfamily: myosin motor domain homology
F;60-727/Domain: myosin motor domain homology <MMO>

Query Match 3.7%; Score 153.5; DB 2; Length 2058;
Best Local Similarity 24.2%; Pred. No. 0.1;
Matches 55; Conservative 40; Mismatches 95; Indels 37; Gaps 7;



Search completed: January 10, 2003, 02:24:48
Job time : 48.9398 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 20:42:30 ; Search time 22.7203 Seconds
(without alignments)
1434.857 Million cell updates/sec

Title: US-09-803-126-6
Perfect score: 4129
Sequence: 1 MYQQRFSFVFVFVQPSPPPK.. .. WPPCSTSARPTGIPPQGRTC 786

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1159.5	28.1	3511	1	MY15_MOUSE	Q9qzz4 mus musculu
2	1155.5	28.0	3530	1	MY15_HUMAN	Q9ukn7 homo sapien
3	409	9.9	2215	1	MY7A_HUMAN	Q13402 homo sapien
4	409	9.9	2215	1	MY7A_MOUSE	P97479 mus musculu
5	192	4.7	1577	1	MYSH_ACACA	P47808 acanthamoeb
6	154.5	3.7	1426	1	CUT2_MOUSE	P70298 mus musculu
7	153.5	3.7	1505	1	CUT2_HUMAN	O14529 homo sapien
8	153.5	3.7	2058	1	MY10_HUMAN	Q9nde7 homo sapien
9	152.5	3.7	2052	1	MY10_BOVIN	P79114 bos taurus
10	143	3.5	559	1	WAS1_HUMAN	Q92558 homo sapien
11	140.5	3.4	282	1	ATP5_HUMAN	Q9y2d1 homo sapien
12	138	3.3	449	1	APG_BRANA	P40603 brassica na
13	134	3.2	1468	1	FMN1_MOUSE	Q05860 mus musculu
14	133	3.2	533	1	PXPR_HUMAN	P28702 homo sapien
15	132	3.2	1206	1	FM14_MOUSE	Q05859 mus musculu
16	126.5	3.1	534	1	APG_ARATH	P40602 arabidopsis
17	126	3.1	512	1	GAG_SMSAV	P03330 simian sarc
18	126	3.1	3421	1	TEGU_HSVEB	P28955 equine herp
19	125.5	3.0	1955	1	PC15_HUMAN	Q96qu1 homo sapien
20	125	3.0	497	1	WAS2_HUMAN	Q9y6w5 homo sapien
21	124.5	3.0	656	1	NF43_HUMAN	Q92570 homo sapien
22	124.5	3.0	1386	1	ZAP3_MOUSE	Q9r0i7 mus musculu
23	124	3.0	1618	1	NEST_HUMAN	P48681 homo sapien
24	123.5	3.0	449	1	G3PT_MOUSE	Q64467 mus musculu
25	123	3.0	505	1	WASL_BOVIN	Q95107 bos taurus
26	123	3.0	3149	1	TEGU_EBV	P41479 autographa
27	122.5	3.0	224	1	Y0V1_NPVAC	Q03209 autographa
28	122	3.0	543	1	VP61_NPVAC	P13983 nicotiana t
29	122	3.0	620	1	EXTN_TOPAC	P05142 mus musculu
30	121.5	2.9	261	1	PRP2_MOUSE	P21997 volvox cart
31	121.5	2.9	485	1	SSGP_VOLCA	Q9fpq6 chlamydomon
32	121.5	2.9	555	1	GP1_CHLRE	Q03211 nicotiana t
33	121	2.9	426	1	EXLP_TOPAC	Q13164 homo sapien
34	121	2.9	815	1	MK07_HUMAN	Q60401 homo sapien
35	120.5	2.9	505	1	WASL_HUMAN	P27460 cas br e mu
36	120.5	2.9	536	1	GAG_MLVCB	Q9jlk4 mus musculu
37	120.5	2.9	1567	1	FMN2_MOUSE	O43516 homo sapien
38	120	2.9	503	1	WAIP_HUMAN	Q9wvs8 mus musculu
39	120	2.9	806	1	MK07_MOUSE	Q9wu42 mus musculu
40	120	2.9	2472	1	NCR2_MOUSE	Q9z1c1 rattus norv
41	119.5	2.9	501	1	WASL_RAT	P46379 homo sapien
42	119.5	2.9	1132	1	BAT3_HUMAN	Q41232 saccharomyc
43	119.5	2.9	1953	1	BN11_YEAST	O60287 homo sapien
44	119.5	2.9	2248	1	Y539_HUMAN	P03332 moloney mur
45	119	2.9	538	1	GAG_MLVMO	

ALIGNMENTS

```
RESULT 1
MY15_MOUSE
ID   MY15_MOUSE     STANDARD;      PRT;  3511 AA.
AC   Q9QZZ4; O70395; Q9QWL6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin XV (Unconventional myosin-15).
GN   MYO15A OR MYO15.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=20021762; PubMed=10552926;
RA   Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
RA   Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
RA   Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,
RA   Fridell R.A.;
RT   "Characterization of the human and mouse unconventional myosin XV
RT   genes responsible for hereditary deafness DFNB3 and shaker 2.";
RL   Genomics 61:243-258(1999).
RN   [2]
RP   SEQUENCE OF 1168-2870 FROM N.A., AND VARIANT SH2 TYR-1779.
RC   TISSUE=Embryo;
RX   MEDLINE=98267310; PubMed=9603735;
RA   Probst F.J., Fridell R.A., Raphael Y., Saunders T.L., Wang A.,
RA   Liang Y., Morell R.J., Touchman J.W., Lyons R.H., Noben-Trauth K.,
RA   Friedman T.B., Camper S.A.;
RT   "Correction of deafness in shaker-2 mice by an unconventional myosin
RT   in a BAC transgene.";
RL   Science 280:1444-1447(1998).
RN   [3]
RP   SEQUENCE OF 1237-1823 FROM N.A., AND VARIANT SH2 TYR-1779.
RC   STRAIN=C57BL/6;
RX   MEDLINE=98369604; PubMed=9703981;
RA   Wakabayashi Y., Takahashi Y., Kikkawa Y., Okano H., Mishima Y.,
RA   Ushiki T., Yonekawa H., Kominami R.;
RT   "A novel type of myosin encoded by the mouse deafness gene shaker-2.";
RL   Biochem. Biophys. Res. Commun. 248:655-659(1998).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC       SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
CC       THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- TISSUE SPECIFICITY: IN THE DEVELOPING INNER EAR, EXPRESSED IN
CC       COCHLEA AND VESTIBULAR APPARATUS. EXPRESSION APPEARS TO BE
CC       RESTRICTED TO COCHLEAR NEUROSENSORY CELLS AND UPPER EPITHELIAL
CC       LAYER OF MACULA SACCULA. ALSO EXPRESSED IN MACULA UTRICULI AND
CC       CRISTAE AMPULLARIS OF THE SEMICIRCULAR CANALS. IN ADULT COCHLEAR
CC       HAIR CELLS, HIGHEST EXPRESSION IN STEREOCILIA AND APICAL BODY.
CC   -!- DISEASE: DEFECTS IN MYO15A ARE THE CAUSE OF SHAKER 2 (SH2), A
CC       CONDITION CAUSING DEAFNESS, CIRCLING BEHAVIOR, HEADTOSSING AND
```

```
CC       HYPERACTIVITY. AUDITORY HAIR CELLS OF AFFECTED ANIMALS HAVE VERY
CC       SHORT STEREOCILIA AND A LONG ACTIN CONTAINING PROTRUSION AT THEIR
CC       BASAL END.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- ----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborat
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ----------------------------------------------------------------------
DR   EMBL; AF144095; AAF05904.1; -.
DR   EMBL; AF054130; AAC40124.1; -.
DR   EMBL; AF031470; BAA36587.1; -.
DR   HSSP; P10587; 1BR2.
DR   MGD; MGI:1261811; Myo15.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00373; myosin_head; 1.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD002356; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS00660; BAND 4.1_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND 4.1_2; FALSE_NEG.
DR   PROSITE; PS50057; BAND 4.1_3; FALSE_NEG.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Disease mutation; Deafness.
FT   DOMAIN          1   1971       HEAD OR MOTOR DOMAIN.
FT   DOMAIN       1872   2013       NECK OR REGULATORY DOMAIN.
FT   DOMAIN       2014   3511       TAIL.
FT   DOMAIN       1307   1334       COILED COIL (POTENTIAL).
FT   DOMAIN       1779   1787       ACTIN BINDING (POTENTIAL).
FT   DOMAIN       1886   1908       IQ 1.
FT   DOMAIN       1909   1938       IQ 2.
FT   DOMAIN       2848   2934       SH3.
FT   DOMAIN       3187   3424       BAND 4.1 LIKE.
FT   NP_BIND      1299   1306       ATP (POTENTIAL).
FT   VARIANT      1779   1779       C -> Y (IN SH2).
FT   CONFLICT     1330   1331       MISSING (IN REF. 2).
FT   CONFLICT     1579   1579       L -> R (IN REF. 3).
FT   CONFLICT     1955   1972       MISSING (IN REF. 2).
FT   CONFLICT     2077   2077       L -> M (IN REF. 2).
FT   CONFLICT     2139   2139       T -> P (IN REF. 2).
FT   CONFLICT     2953   2953       V -> A (IN REF. 2).
SQ   SEQUENCE   3511 AA,  385233 MW,  3B2D2E32A2153B6B CRC64;

  Query Match             28.1%;  Score 11..5;  DB 1;  Length 3511;
  Best Local Similarity   34.3%;  Pred. No. 9.1e-65;
  Matches  845;  Conservative  130;  Mismatches  245;  Indels  173;  Gaps   22;

Qy        4 SRPGPVPVPVQPSPPPPAPIPPIPPPDPALAPLPGINPAHSPPPMLSPSPGPGPPPAVAPP 63
            :|.     :: |   || :                 |  : | :: |:|    | :||
Db     2406 SPPQLPPPILAPLARPPIVQ                ISPPAPPVLVPPTPISWAPGSVAPA 2513

Qy       64 PKAP----------LQLGPS-------------------SSIKE-PQGP-- 93
            :.                :  |                   | ||: || |
Db     2514 PRIPSEPVAVTILAGPWTAPEGPAGPPGPPGPPGPPGPPGPPGPPGPPGPWA 2567

Qy       83 ------------------------------------------------------ 138
                                                    |: | ||:  :| |  :  :     : | | :| :
Db     2511 GPPPAPPPPGPVPPPPGPPPPPPLPAPMPPPGPGPAVVPPGPVPPPAVAMVPPVT 2566

Qy      139 RGL-          PGPVPEPGVPLPLAPSGPVPPPPPTTWKLPLAKPVPVPPPPNPGHP 188
             |:           :| || | |||          : |   :|: :: :|:|:|:|:|:
Db     2567 PCMGPPPVPLPPSLLPPLLP VGTGLPHPLVNPNPPVGPLPTWKPTLPPPVPPPPLNYGHP 2626

Qy      189 YYPPLPPGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 248
             | ||  ||| ||||||:|:||||:|| :|| |    ::|      :: ||||:  : ||
Db     2627 VGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2686

Qy      249 PPWANYPGPPPPVSGPGPGPPPQPLAVGHPGLPLPVTGPPPPPPPLPPPPYGPPAPVLP 308
             |:|  |||| ||  |  :| ||:||||| |::||:: :|      :|::||:|||':|
Db     2687 PGWPIYPSPLPPAMGSVGTPVQPLAVGHTGIPLLGMVPPPPPPAPPPPPVLPAYGPAPLLP 2746

Qy      309 VEGPGGSTLELSLKSSGQLVLHTAPAPAIEALVELPLNELPPLGPPPCAPAPYLPGPGLL 368
             |    : || :| :|:|:| :|||: :: || |: |:|||| '|:|:::::::  |:
Db     2747 VTMPPPGMVPPPNPPPLPPPAPAGLVPTTVPTPPPPPPPLSGYVVAVPNPLGEDPELL 2806

Qy      369 SPHPGDLPLLLPVATLEP- -- - - - - - - - - - - - - - - -GWQPGSAPGPGPGLP 407
             |||:|||:| |    :|||                              |:::|:  |: | |
Db     2807 SPHPGDLPLLLP  GGLPPTAPGPGAGGYPPPPLVTLPPLPPPPGPGPPWPPGAVHPPVGPGP 2866

Qy      408 PADIVQPAAAPDPSPSKEQPPGWHKGQLSNGEPGLAR - - - - - WDPAGPPPAH 457
             |:::|||||||||             || |:|| |                  :
Db     2867 PSPLVQPAAAPDPL            GLP AEPGPGPAAAVVAAVAPAAAAGEVGPPPPG 2921

Qy      458 PWSQAHSDDSEATSLSSVAYAFLPDSHSYTMQEFARRYFPRRG      ALLQQTPPAAG 507
             |  :| | ||   |::       || |   || ||:|:||:|:  :      :  ::
Db     2922 PPVPAPSAPGGPPGIA  - IPPP  TMPPPA PYPPPPPPPPPPPPPPPPPPPPPPG 2977

Qy      508 KLTLSLVQYTPAPLQESLSLSLSLSPVPLAVAPPLALMPPMPPPLPPPPPPPPPLPPP 567
             |  : :: :|| |||||: ||| :::|:||  |:|:||'':  : :: ':|:|  ||
Db     2978 KTLEPVLPPKVPTQESLIELSPNLNPMAVPMPPVAVMPPMPPA APPPPPPPPPPTLL 3037

Qy      568 KLG-QQEKLPDEIYQQVIKQVTQH-PPPPHCTPGWSPPPPPPPPPPPPPPPPPPPPP 626
             |||   | :||| |||:|||:| :  |: : | |||  | ::  ::  |   ||| :||
Db     3038 KLPPPHVMPPPVYQPVPKLIINPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 3097

Qy      627 QPPP------PSQPLAPPSQPPHQPVPVPPPPMPPPPPPMPPAPPPPQAIPLLLIPPPPVV 680
             |  | :|:: ::|||:|:::|||   |  |:|: | |  :  :  |  ||| :|:|:
Db     3147 QPVPWTPPPPPPQTIAPAPPQNLPPIPPPPPPPPPPPPPPMPPPAMPAPPPPPPPPPPPP 3206

Qy      681 DYRTNIQTFTVAAEVQEELGRQMGITEPQRVQRFALPLIPEESQQLVRPLPAEYLNQVVV 740
             :   |:| ||| :| | || :| :| |:   |: :|::  : |  ||    |: |
Db     3207 PPHLPIPTPTVAIPVIPGLTPPMALTPPPAPPTPVIPTTTPPPQPPPPPPPPAAPPPPP 3266

Qy      741 DQD 743
             : :
Db     3267 EME 3269


RESULT 2
MY15_HUMAN
ID   MY15_HUMAN     STANDARD,      PRT;  3530 AA.
AC   Q9UKN7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin XV (Unconventional myosin-15).
GN   MYO15A OR MY15.
OS   Homo sapiens (Human).
OG   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., CHROMOSOMAL LOCATION, AND TISSUE SPECIFICITY.
RX   MEDLINE=20021762; PubMed=10552926;
RA   Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
RA   Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
RA   Sullivan K.F., Tepass U., Riazuddin S., Kachar B., Friedman T.B.,
RA   Fridell R.A.;
```

```
RT   "Characterization of the human and mouse unconventional myosin XV
RT   genes responsible for hereditary deafness DFNB3 and shaker 2.".
RL   Genomics 61:243-258(1999).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A., AND VARIANTS DFNB3 TYR 2111 AND PHE-2113.
RX   MEDLINE=98267311; PubMed=9603736;
RA   Wang A., Liang Y., Fridell R.A., Probst F.J., Wilcox E.R.,
RA   Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K., Camper S.A.,
RA   Friedman T.B.;
RT   "Association of unconventional myosin MYO15 mutations with human
RT   nonsyndromic deafness DFNB3.";
RL   Science 280:1447-1451(1998).
CC   -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC       SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
CC       THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY. ALSO EXPRESSED
CC       AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,
CC       PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN. IN THE
CC       PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.
CC   -!- DISEASE: Defects in MYO15A are the cause of an autosomal recessive
CC       form of nonsyndromic deafness (DFNB3).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 PH DOMAINS.
CC   ---------- . ---------- . . . . . . . . . . . . . . . . . . . . . . . . . .
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licenses@isb-sib.ch)
CC   ---------- . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
DR   EMBL; AF144094; AAF05903.1; -.
DR   EMBL; AF051976; -; NOT_ANNOTATED_CDS.
DR   HSSP; P10587; 1BR2.
DR   Genew; HGNC:7594; MYO15A.
DR   MIM; 602666; -.
DR   MIM; 600316; -.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_motif_EF-hand.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000455; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00822; BAND_4_1_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_4_1_2; FALSE_NEG.
DR   PROSITE; PS00657; BAND_4_1_3; 1.
DR   PROSITE; PS50096; IQ; 3.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Calmodulin binding; Disease mutation; Deafness.
FT   DOMAIN          1   1887       HEAD OR MOTOR DOMAIN.
FT   DOMAIN       1888   2029       NECK OR REGULATORY DOMAIN.
FT   DOMAIN       2030   3530       TAIL.
FT   DOMAIN       1323   1350       COILED COIL (POTENTIAL).
FT   DOMAIN       1792   1799       ACTIN-BINDING (POTENTIAL).
FT   DOMAIN       1902   1924       IQ 1.
FT   DOMAIN       1925   1954       IQ 2.
FT   DOMAIN       1955   1976       IQ 3.
FT   DOMAIN       2867   2953       SH3.
FT   DOMAIN       3206   3443       BAND 4.1-LIKE.
FT   NP_BIND      1315   1322       ATP (POTENTIAL).
FT   VARIANT      2111   2111       N -> Y (IN DFNB3; FAMILY FROM BENGFALA).
FT                                  /FTId=VAR_010303.
FT   VARIANT      2113   2113       I -> F (IN DFNB3; INDIAN FAMILY).
FT                                  /FTId=VAR_010304.
SQ   SEQUENCE   3530 AA;  395171 MW;  3D103823D4BCBE4A CRC64;

  Query Match           28.0%;  Score 1155.5;  DB 1;  Length 3530;
  Best Local Similarity 35.9%;  Pred. No. 1.6e-64;
  Matches  277,  Conservative  131;  Mismatches  223;  Indels  141;  Gaps   20;

Qy      12 PVQPSRPP-------KAFLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRP 64
           |  |  |       | |.:: || .|||  |                        ||    :| |
Db    2586 PFHGGPPEALPKDGGKVFMKRPDPHEEALMIL-------- .  . . -KGQMTHLAAAP 2628

Qy      65 KAPLQLGPSSSIKPPQGPLLDLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLTQP 124
           |      :  . :|| |                |:   | |  ||   ||        | | |
Db    2629 GTQVSREAVALVK---------- . PVTSAPPPSMAPISALPS-       . RSLPPP 2665

Qy     125 VR-DQGVSTQLLAPSGSVCPSYTGTPWKLPLRKEVPYPPSNFNSPFTYTSLLCPQILPPTP 183
           |  |  :|. |:   : | |||:|||||||||:|||  | ||  ||| ||
Db    2666 EELTQTRLHRLINEN --FYGYQQAIWKIPLRKEVPYPPTSYSHFVSIPSLPPQILHDTL 2722

Qy     184 SEGCIRISQNERKMKDLLGGLEVDLDSLTTEDSVKKRIVVAARDNWANYFSPFFPVSG 243
           |:|:|||:.|| :|| |  :::     || |||. :|  ||| | |||| || :|
Db    2723 SEACLPISEDEKLRMKALPAQNGLDTQFPLVTF-SVKPAVVSTAFDTWEVYFSKIFPATG 2781

Qy     244 ESGSLVQLLAVSHRGLPLLFVTQGPGLRPDQLKILCSYSFAEVLGVECRGGSTLELSLKS 303
           |. |||||| |..:|:.. .|     |||:| :||||::| | :  :|| :| |
Db    2782 SVGTGVQLLAVSHVGIKLLRMVKGGQEAGGQLRVLRAYSFADILPVTMPSQNMLEFNLAS 2841

Qy     304 EQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHPGDLIKLLPVAT 363
           |...| :|||  :: || |: ||||| ||.|.|::: :: :|||:||:|| | |
Db    2842 EEVIIPSAPAHQVFTLVDDFILELKKDSDYVVAVRNFLPEDPALLAFHKGDIIHIQD  2898

Qy     364 LEP----------------  ---  .  . GWQFGSAGGRSGLFPADIVQIAAAPPFGP 395
           |||                                ||:||:  || | ||:::|||||||||
Db    2899 LEPPRVGYSAGCVVRRKVVYLEELRRRGPDFGWRFGTIHGRVGRFPSELVQPAAAPDFL  2957

Qy     396 SKEQRSGWHKGQLSNGEPGLAR---------- .  . . .-WDPASEPPAHPWSQAHSDDSE 438
           |  |  |||  |                                |  |  |  | :|:
Db    2958 ----------QLPT-EPGRGRAAAVAAAVASAAAAQEVGPEEEGPDVPAPSADHGPDAL 3005

Qy     439 ATSLSSVAYAFLPDSHSYTMQEFAPPYFPPSQ-----ALLGQTDGGAAGKDTDSLVQVTK 493
           |              ||   ||| |||:|||  |     |  ::          : :  :||
Db    3006 A----------LP---PYTMLEFAQEYFPDPQFPPQPDGLPLKSKEPPPSPTLEDMLCFTK 4052

Qy     494 APIQESLLSLSD-DVSKLAVASFLAIMRFMGDQSKPPGKDEMDLLYELLKLC-QQEKLPD 551
           |.|||:  ||  :|:|:   ||| |||||| :  .|. ::|:| |||||  | :|'
Db    3053 TPLQESLIELSDSSSLSKMATDMFLAVMRFMGD-APLPGQSCLPVLCNLLKLCGGDHFVMRD 3111

Qy     552 EIYCQVIKQVTGH--PPPEHCTRGWSFLSLLTGFFPPSTRLMFYLTKFLQDSG-----PS 604
           | |||:||:| :  : : |||  | ::|  :   | |:|:||||       |
Db    3112 ECYCQVVKQITDNTSSKQDSVCPRWRLLYIVTAYHSCSEVLHPHLTRFLQDVSRTFGLDF 3171

Qy     605 QELAKSSQEHLQKTVKYGSFPMPPPGRMKAFLKGQAIPLLLIHLPGGVDYRTNIQTFTV 664
           |  |:  :::|:|: :|||  :|    :  | | |:: :  |  ||||:    |:| ||
Db    3172 QGIAFAGEQNLQKTLRFGGRLELPSSIELRAMLAGRSSKRQLFLDLSGLERHLFTKTGTV 3231

Qy     665 AAEVQEELCKQMGITEFQEVQEFALFLIKEKSQLVPPLQFAEYLNGVVVDQD 716
           | :|  |:| :| :| |.   |. :|:.  . | | |:    :  | :
Db    3232 ALDVVEFICAFMALTRPEAFNEYVIFVVTNRGQHVCFLSFRAYILDVAGEME 3283


RESULT 3
MY7A_HUMAN
ID   MY7A_HUMAN     STANDARD;      PRT;  2215 AA.
AC   Q13402; Q96R21; Q96R22; Q13331; P78427; Q14785;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

DE Myosin VIIa.
GN MYO7A OR USH1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1, 5, 6 AND 7), AND DEVELOPMENTAL
RP EXPRESSION.
RC TISSUE=Retina;
RX MEDLINE=96194648; PubMed=8600819;
RA Weil D., Levy G., Sahly I., Levi-Acobas F., Blanchard S.,
RA El Amraoui A., Crozet F., Philippe H., Abitbol M., Petit C.;
RT "Human myosin VIIA responsible for the Usher 1B syndrome: a predicted
RT membrane associated motor protein expressed in developing sensory
RT epithelia.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3232-3237(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 4).
RC TISSUE=Testis;
RX MEDLINE=97218686; PubMed=8884267;
RA Chen Z.-Y., Hasson T., Kelley P.M., Schwender B.J., Schwartz M.F.,
RA Ramakrishnan M., Kimberling W.J., Mooseker M.S., Corey D.P.;
RT "Molecular cloning and domain structure of human myosin VIIa, the gene
RT product defective in usher syndrome 1B.";
RL Genomics 36:440-448(1996).
RN [3]
RP SEQUENCE OF 166-196 FROM N.A.
RC TISSUE=Epithelium, Leukocyte, and Liver;
RX MEDLINE=94294418; PubMed=8022818;
RA Bement W.M., Hasson T., Wirth J.A., Cheney R.E., Mooseker M.S.;
RT "Identification and overlapping expression of multiple unconventional
RT myosin genes in vertebrate cell types.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:6549-6553(1994).
RN [4]
RP SEQUENCE OF 1-1075 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=96120870; PubMed=7608974;
RA Hasson T., Heintzelman M.B., Santos-Sacchi J., Corey D.P.,
RA Mooseker M.S.;
RT "Expression in cochlea and retina of myosin VIIA, the gene product
RT defective in Usher syndrome type 1B.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9815-9819(1995).
RN [5]
RP SEQUENCE OF 96-564 FROM N.A. (ISOFORM 1), AND VARIANTS USH1B.
RC TISSUE=Retina;
RX MEDLINE=95174880; PubMed=7870171;
RA Weil D., Blanchard S., Kaplan J., Guilford P., Gibson F., Walsh J.,
RA Mburu P., Varela A., Levilliers J., Weston M.D., Kelley P.M.,
RA Kimberling W.J., Wagenaar M., Levi-Acobas F., Larget-Piet D.,
RA Munnich A., Steel K.P., Brown S.D.M., Petit C.;
RT "Defective myosin VIIA gene responsible for Usher syndrome type 1B ",
RL Nature 374:60-61(1995).
RN [6]
RP SEQUENCE OF 79-578 FROM N.A.
RX MEDLINE=97224487; PubMed=9070921;
RA Kelley P.M., Weston M.D., Chen Z.-Y., Orten D.J., Hasson T.,
RA Overbeck L.D., Pinnt J., Talmadge C.B., Ing P., Mooseker M.S.,
RA Corey D.P., Sumegi J., Kimberling W.J.;
RT "The genomic structure of the gene defective in Usher syndrome type 1b
RT (MYO7A).";
RL Genomics 40:73-79(1997).
RN [7]
RP DEVELOPMENTAL EXPRESSION.
RX MEDLINE=96440426; PubMed=8842737;
RA El Amraoui A., Sahly I., Picaud S., Sahel J., Abitbol M., Petit C.;
RT "Human Usher 1B/mouse shaker-1: the retinal phenotype discrepancy
RT explained by the presence/absence of myosin VIIA in the photoreceptor
RT cells.";
RL Hum. Mol. Genet. 5:1171-1178(1996).
RN [8]
RP VARIANTS USH1B ARG-25, SER-955 AND GLU-2147, AND POLYMORPHISMS.
RX MEDLINE=97162218; PubMed=9007678;
RA Levy G., Levi-Acobas F., Blanchard S., Gerber S., Larget-Piet D.,
RA Chouai M., Liu X.-Z., Newton V., Steel K.P., Brown S.D.M., Munnich A.,
RA Kaplan J., Petit C., Weil D.;
RT "Myosin VIIA gene: heterogeneity of the mutations responsible for
RT Usher syndrome type IB.";
RL Hum. Mol. Genet. 6:111-116(1997).
RN [11]
RP VARIANT DFNB2 PRO-244.
RX MEDLINE=97315827; PubMed=9171832;
RA Liu X.-Z., Walsh J., Mburu P., Kendrick-Jones J., Cope M.J.,
RA Steel K.P., Brown S.D.M.;
RT "Mutations in the myosin VIIA gene cause non-syndromic recessive
RT deafness.";
RL Nat. Genet. 16:188-190(1997).
RN [12]
RP VARIANT DFNB2 ILE-599.
RX MEDLINE=97315828; PubMed=9171833;
RA Weil D., Kussel P., Blanchard S., Levy G., Levi-Acobas F., Drira M.,
RA Ayadi H., Petit C.;
RT "The autosomal recessive isolated deafness, DFNB2, and the Usher 1B
RT syndrome are allelic defects of the myosin VIIA gene.";
RL Nat. Genet. 16:191-193(1997).
RN [13]
RP VARIANT DFNA11 ALA-886 ALA-888 DEL.
RX MEDLINE=98016404; PubMed=9354784;
RA Liu X.-Z., Walsh J., Tamagawa Y., Kitamura K., Nishizawa M.,
RA Steel K.P., Brown S.D.M.;
RT "Autosomal dominant non-syndromic deafness caused by a mutation in the
RT myosin VIIA gene.";
RL Nat. Genet. 17:268-269(1997).
RN [14]
RP VARIANTS USH1B PRO-681 AND GLN-1602.
RX MEDLINE=98186030; PubMed=9718356;
RA Liu X.-Z., Hope C., Walsh J., Newton V., Ke X.-M., Liang C.-Y., Xu L.-R.,
RA Chen J.-M., Tsang T., Steel K.P., Bradley D., Brown S.D.M.;
RT "Mutations in the myosin VIIA gene cause a wide phenotypic spectrum,
RT including atypical Usher syndrome.";
RL Am. J. Hum. Genet. 63:909-912(1998).
RN [15]
RP VARIANT USH1B PRO-1087.
RX MEDLINE=99244608; PubMed=10364543;
RA Adato A., Kalinski H., Weil D., Chaib H., Korostishevsky M.,
RA Bonne-Tamir B.;
RT "Possible interaction between USH1B and USH3 gene products as implied
RT by apparent digenic deafness inheritance.";
RL Am. J. Hum. Genet. 65:261-265(1999).
RN [16]
RP VARIANTS USH1B, AND POLYMORPHISMS.
RX MEDLINE=99145257; PubMed=10094549;
RA Janecke A.R., Mullis M., Steinlein O.K., Grunt T.W., Agha-Zadeh F.,
RA Tyler E., Gottlob I., Renner F., Rosenberg T., Gal A.;
RT "Twelve novel myosin VIIA mutations in 34 patients with Usher syndrome
RT type I: confirmation of genetic heterogeneity.";
RL Hum. Mutat. 13:133-140(1999).
RN [17]
RP VARIANTS USH1B ILE-1178 AND TYR-1719.

```
RX   MEDLINE=99374392; PubMed=10447283;
RA   Cuevas J.M., Espinos C., Millan J.M., Sanchez F., Trujillo M.J.,
RA   Ayuso C., Beneyto M., Najera C.;
RT   "Identification of three novel mutations in the MYO7A gene.";
RL   Hum. Mutat. 14:181-181(1999).
CC   -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN
CC       RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RIBBON-
CC       SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTORS
CC       DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE RIGIDITY OF STEREOCILIA
CC       DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE. INVOLVED IN HAIR-CELL
CC       VESICLE TRAFFICKING OF AMINOGLYCOSIDES, WHICH ARE KNOWN TO INDUCE
CC       OTOTOXICITY (BY SIMILARITY).
CC   -!- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE THROUGH THE
CC       FORMATION OF A COILED-COIL ROD.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE). IN THE PHOTORECEPTOR
CC       CELLS, MAINLY LOCALIZED IN THE INNER AND BASE OF OUTER SEGMENTS AS
CC       WELL AS IN THE SYNAPTIC ENDING REGION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; 1 (SHOWN HERE), 2, 3,
CC       4, 5, 6 AND 7, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE PIGMENT EPITHELIUM AND THE
CC       PHOTORECEPTOR CELLS OF THE RETINA. ALSO FOUND IN KIDNEY, LIVER,
CC       TESTIS, COCHLEA, LYMPHOCYTES. NOT EXPRESSED IN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: DETECTED IN OPTIC CUP IN 5.5 WEEKS-OLD
CC       EMBRYOS. EXPRESSED IN RETINAL PIGMENT EPITHELIUM, COCHLEAR AND
CC       VESTIBULAR NEUROEPITHELIA, AND OLFACTORY EPITHELIUM AT 8 WEEKS. AT
CC       10 WEEKS, PRESENT IN BOTH PIGMENT EPITHELIUM AND PHOTORECEPTOR
CC       CELLS. AT 24-24 WEEKS, EXPRESSION IN PIGMENT EPITHELIUM AND
CC       PHOTORECEPTOR CELLS INCREASES, PRESENT IN PIGMENT EPITHELIUM AND
CC       PHOTORECEPTOR CELLS IN ADULT.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF USHER SYNDROME TYPE 1B
CC       (USH1B), ALSO KNOWN AS USHER SYNDROME NON-ACADIAN VARIETY. USHER
CC       SYNDROME TYPE 1 IS AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY
CC       PROFOUND CONGENITAL SENSORINEURAL DEAFNESS, CONSTANT VESTIBULAR
CC       DYSFUNCTION AND PREPUBERTAL ONSET OF PROGRESSIVE RETINITIS
CC       PIGMENTOSA LEADING TO BLINDNESS. USHER SYNDROME IS THE MOST COMMON
CC       CAUSE OF COMBINED DEAFNESS AND BLINDNESS IN DEVELOPED COUNTRIES.
CC       USHB1 APPEARS AS A PRIMARY CYTOSKELETAL PROTEIN DEFECT.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB2 OR NSPD2) WITH
CC       VARIABLE VESTIBULAR DYSFUNCTION AND VARIABLE AGE OF ONSET.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA11) WITH ONSET
CC       AFTER COMPLETE SPEECH ACQUISITION AND SUBSEQUENT GRADUAL
CC       PROGRESSION.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 IQ DOMAINS.

  Query Match              9.9%;  Score 409;  DB 1;  Length 2215;
  Best Local Similarity   23.1%;  Pred. No. 9.4e-18;
  Matches  189;  Conservative  111;  Mismatches  264;  Indels  254;  Gaps   30;

Qy      30 DEALAKLGINGAHSSPFM--LSPSPGKGP---PPAVAPRPKAPLQLGPSSSIKE----- 78
           :| |  :  :|| : ||    |  : | |    |          |     :: ||
Db    1228 EERLRRTFVNGTRTQPFSWLELQAIKSKKPIMLPVTFMDGTTYTLLTDSATTAKELCNAL 1287

Qy      79 -PQGPLLPLPGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRCLTQPVEDQGVSTQLLAP 137
            :  | | ||  | ||                  |  |  |                    .
Db    1288 ADFISLKDFFGFSLYIA------ -- ----------------LFDKVSSLGSGSDHVMD 1321

Qy     138 SGSVCFSTT--- --- GTPWKLFLPKEVFYPRENFSHP-YYLRLLCEQILRDTFSESCI 188
           :  | |  |          |:|| |||| |  |  |  |  |:  |  |  |  .
Db    1322 AIGQCDLQAFEDGALPFNAFWPLFPPKEVFTPWHCPSEDNVATNLIYQQVVPG------V 1375

Qy     189 RISQHERRKMKDLIG- --------GLEVDLDSL------------------------- 212
           :  :   | ||                | |  | |
Db    1376 KFGEYRCEKEDDLAELACQQYFVDYGSEMIIPRLLNLVPTYIPDPEITPLKTLEKWAQLA 1435

Qy     213 -- ---------TTTEDSVKKFIVVAARDNWANYFSPPFPVSGESG-----SDVQLLAVS 255
               |         ||: :|  ||  |  |  |||:     ||     .|| ..|:
Db    1436 IAAHKKGIYACPFTPACPVKEDVVSYAPFFWPLIFSPPYEAYKFSGPSLPKNDV-IVAVN 1494

Qy     256 HRGLPLLYVTQGPGLRPDQLKILCSYSFAEVLGV---------------------------- 289
           |. .                 :  :  :|   |  |
Db    1495 WTGVYPV---------DEQEQVLLELSFPEIMAVSSSPEQPVWLSLGCSLDGCAAHGGW 1545

Qy     290 -------------ECRGGSTLELS-----LKSEQLVLHTARARAIEALVELPLNELKKDS 331
                       |||  |   |     :| ::    :: |  |  ||  ||  |:| |
Db    1546 AGLTTAGPCSPCWSCPGAKTTAPSFTLATIRGLEYTFISSNARDCFPLVYTFLPCLPKPS 1605

Qy     332 GYVTALRSYITDN-------CSLLSFHRGDLIKLLPVATLEPGWQFGSAGG---    RSG 378
           :|.||.    ||          ||| :|||| :|  |  |      | | |      : |
Db    1606 KYVVALQ- - DNPNPAGEESGFLSFARGDLI-ILDHDTGEQVWNSTWANTNEKTFQRG 1660

Qy     379 LFPAD--IVQPAA--------- ---ADDFSFSKFQPSGWHFGQLSNGEPGLAKWDRASE 423
           ||| |  | ||                ||    :::     ||  ||          |
Db    1661 DFPTDCVYVMPTVTMPPKEIVALVTMTPD- - -QRQEVVELILPTAEP---------E 1706

Qy     424 PPAHPWSQAHSDDSEATSLSSVAYAFLPDSHSTTMQEFARSYPF  ----PSQAILGQTD 477
           | |                          ||  :||:  |||          |: : :
Db    1707 VPAKP--------------------------YTLEEFSYDYFRFTFPHTLQPVMVSF-  1737

Qy     478 GGAAGKDTDSLVQYTKAPIQESLLSL---SDDVSKLAVASFLALMKFMGDQSNFRGFTPM 534
           | ||  | |  -|- | :.:|     |.. | |  | :|:|---:|||    | :
Db    1738 --ARGK--DFLWSHTFEPLKQALLKKLIGSEELSQEACLAFIAVLKYMGLYPPFRTRQVN 1793

Qy     535 DLLYELLF-DQQQEKLRDEIYCQVIKQVTGHPFPEHCTRGWSFLSLLTGFFFSCTELMPV 593
           .|   :  :  . | |:|| | |:.||:|  :      ||| |  | ||| ||||  |:|:
Db    1794 ELTDQIPEGPLKAEPLKDEAYVQILKQLTDNHIRYSEERGWELLWDCIGLFFFSGTILLPV 1853

Qy     594 LTKFLQDSGFSQELARSSQEHLQPTVFGGFFRMFPFGEMFAFLFCQAIPLLLTHLPGGV 653
           : :||| |     ||    : ||: .. | |:  |  |::|              :: |
Db    1854 VQRFLQ-SKKHCPLAIDCLQKLGKALENNSEFYFPHLVEVPAIQHPTTQTFRKVYFPDDT 1912

Qy     654 DYRTNIQTPTVAAEVQEELCRQMGITPEQEVQEFALFL 691
           |     :  |  .  :  :   :: :   :  : |:||:
Db    1913 DEAFEVESSTKARDFCQNIATPLLL--KDSSGFSLFV 1947


RESULT 4
MY7A_MOUSE
ID   MY7A_MOUSE     STANDARD;      PRT,  2215 AA.
AC   P97479;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin VIIa.
GN   MYO7A OR MYO7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria; Rodentia; Sciurognathi; Muridae, Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95174861, PubMed=7870172;
RA   Gibson F., Walsh J., Mburu P., Varela A., Brown K.A., Antonio M.,
RA   Beisel K.W., Steel K.P., Brown S.D.M.;
RT   "A type VII myosin encoded by the mouse deafness gene shaker-1.";
RL   Nature 374:62-64(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98343888, PubMed=9680294;
RA   Mburu P., Liu X. Z., Walsh J., Saw D. Jr., Cope M.J., Gibson F.,
RA   Kendrick-Jones J., Steel K.P., Brown S.D.M.;
RT   "Mutation analysis of the mouse myosin VIIA deafness gene.";
RL   Genes Funct. 1:193-203(1997).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN
CC       RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RIBBON-
CC       SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTORS
CC       DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE RIGIDITY OF STEREOCILIA
CC       DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE (BY SIMILARITY).
```

```
CC       INVOLVED IN HAIR CELL VESICLE TRAFFICKING OF AMINOGLYCOSIDES,
CC       WHICH ARE KNOWN TO INDUCE OTOTOXICITY.
CC     - SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE THROUGH THE
CC       FORMATION OF A COILED COIL ROD (BY SIMILARITY).
CC     - SUBCELLULAR LOCATION: Cytoplasm. (Probable).
CC     - TISSUE SPECIFICITY: EXPRESSED IN THE RETINA, COCHLEA, KIDNEY AND
CC       LIVER.
CC     - DEVELOPMENTAL STAGE: IN THE INNER EAR OF THE 16.5 DAY OLD EMBRYO,
CC       EXPRESSED ONLY IN THE COCHLEAR AND VESTIBULAR SENSORY HAIR CELLS.
CC       IN ADDITION, EXPRESSION ALSO OCCURS IN THE EPITHELIAL CELLS OF THE
CC       SMALL INTESTINE, RETINAL CELLS, AND THE FETAL TESTIS.
CC     - DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF THE SHAKER-1 (SH-1)
CC       PHENOTYPE WHICH AFFECTS ONLY THE INNER EAR. SH-1 HOMOZYGOTE
CC       MUTANTS SHOW HYPERACTIVITY, HEAD TOSSING AND CIRCLING DUE TO
CC       VESTIBULAR DYSFUNCTION, TOGETHER WITH TYPICAL NEUROEPITHELIAL-TYPE
CC       COCHLEAR DEFECTS INVOLVING DYSFUNCTION AND PROGRESSIVE
CC       DEGENERATION OF THE ORGAN OF CORTI.
CC     - SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC     - SIMILARITY: CONTAINS 5 IQ DOMAINS.
CC     - SIMILARITY: CONTAINS 2 BAND 4.1 LIKE DOMAINS.
CC     - SIMILARITY: CONTAINS A SH3 DOMAIN.
CC       -!- ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC       -!- ------------------------------------------------------------
DR   EMBL; U81458; AAB40708.1; -.
DR   HSSP; P08799; 1MND.
DR   MGD; MGI:104510; Myo7a.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR...; MYOSINHEAVY.
DR   ProDom; PD000066; SH3; 1.
DR   ProDom; PD...; ...band_4.1; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00233; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; ...; BAND_4_1_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_4_1_2; FALSE_NEG.
DR   PROSITE; PS00657; BAND_4_1_3; 2.
DR   PROSITE; PS50096; IQ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Myosin; ATP binding; Actin binding; Coiled coil; Repeat; SH3 domain;
KW   Calmodulin binding; Disease mutation; Deafness.
FT   DOMAIN          1    729       MYOSIN HEAD-LIKE.
FT   DOMAIN        745    765       IQ 1.
FT   DOMAIN        768    788       IQ 2.
FT   DOMAIN        791    811       IQ 3.
FT   DOMAIN        814    834       IQ 4.
FT   DOMAIN        837    857       IQ 5.
FT   DOMAIN        908    935       COILED COIL (PROBABLE).
FT   DOMAIN       1255   1441       BAND 4.1 LIKE 1.
FT   DOMAIN       1623   1672       SH3.
FT   DOMAIN       1809   2139       BAND 4.1 LIKE 2.
FT   NP_BIND       149    165       ATP (PROBABLE).
FT   DOMAIN        572    639       ACTIN BINDING (PROBABLE).
FT   VARIANT       241    241       R -> P (IN SH-1).
FT   VARIANT       502    502       R -> P (IN SH-1).
SQ   SEQUENCE   2215 AA;  251847 MW;  3470CEL1941ABLED CRC64;
```

```
  Query Match              9.4%;  Score 409;  DB 1;  Length 2215;
  Best Local Similarity   27.6%;  Pred. No. 9.4e-19;
  Matches  190;  Conservative  114;  Mismatches  241;  Indels  154;  Gaps   9;

Qy      36 PFAIAPLSINEAMSIDPM  LIPCFPSP  PPAVAPPEPAPLSSHVSSIEE      79
           :| | :  :||  : ||    |  :  | |   |                :  ::: :|
Db    1228 EERLRKIPVNGIRQPPSWLEIQAIRSKRPIMDPVIPMD-IKI...SAIIAPEICNAL 1287

Qy      80 PLSPLLLLPSIPLLIAHIIIIIAFILLLLL CNLSAEPRGLTSIVEPSGVSTSLLAP 137
           :  | | || | ||'                                :  |  | : :
Db    1288 APPICIPIPPDPIIYIA                              ...PS. . GTTSDPM 1331

Qy     138 SSVSPLYT        STPWPLILPPPLPTPPPMLSP ::::::. P.SLPL::::. I 188
           :  |  |              |.:|| |||||  |  |  :       :  :  |  :  :
Db    1332 AI.P.PAPIPMAPPPAPWPIIPPPIPIPIWHPP PIPLAIL:::...VPP       V 1378

Qy     189 PIPQNPPPPMYPLIP          GIPVDLDSL                          3?
           :  :    |  ||              | |: |: |
Db    1379 PPSPSPLPLLLLAPLA.P.PPVLIS.PMILPPI...IVI:::::P:::::PNLPPWAPLA 1438

Qy     239              TTIPLDVPPPITVAAPLNWADYPIPPIPVPPPP       SDVPLLAVS 285
                        |    ||: :|  :|  |  :|:|:          :      |  :::::
Db    1439 IAAHPPSIYAPPPTPPIPVPIITVDVAPPPWPIPPPPITAVPPP:::::PPPIN IVAVN 1496

Qy     286 HPGLPLLPVTQSPGLPPDQLPTLCSYSPAEVIAIV  ::PPR           PGSTLP    338
           ::  :          :|  ::|   || |:: |     ||:           |  : |
Db    1497 WTPVTPV         PPPPLPLLPIPIMAVPSPPPPVLIPLIIIIPVPVPPPPPH 1540

Qy     339                    IPLPPPPQLVLHTAIPVPVPAIVPPIIPPIPPPP 391
                              ::|  ::    :: :    :  :|  ||  |  |
Db    1541 AGLPPAGPPSPPWSPPGIPMMAPPPIIATIPGPPYTPTPSNARPIPPLVVTPLPPLPPPPS 1600

Qy     392 GTVIALPPSYITPN-------QSLISPPPPPPIPLIPVATLPPPWPPSAGS      PPP 438
           |:|||:    ||          ||| :|||| :|                        : |
Db    1601 KYVVALP----PNPNPAGPPPSPLPPAPGDLI LIPHPTPPPPVMNTPWAPIPPPTPLPP 1655

Qy     439 LPPAL--IVQPAA-------------APPPPPPPPPPPPPPPPPPPPPPPPPPPPP 478
           ||  |   |                 ||     ::    :    :|  ::
Db    1656 DPPTPPCVYVMPTVTLPPPPEIVALVTMPPP     QPQPVVPLLPPTAPI          E 1704

Qy     479 PPAHPWPQAHSDDSEATSLSSVAYAPLPPSHSYTMPPAPPYPP       PPPPPPPPP 527
           |||                                 ||:|||:  |||              :
Db    1705 VPAPI                               YTLPPPPSTPYPPPPPPPPTLPPVVVPP   1747

Qy     528 PGAAPPYPPPPPYQYPPAPPQPPPPP    PPPPPPPPPAPPPPPPPPPPPPPPPPPPP 576
           | ||  | |  :|: |:::||     |:::|: |  :::::::::  :    | :
Db    1748 --ARGK--DRLWSHTPPPLPQAILKKILGSEELSQPAPMAPVAVLKYMGDYPPKPMPVVN 1797

Qy     577 PLLYPLLP PPPPPPLPPPPYPPQVPPPPVPPPPPPPPPPPPPPPPPPPPPPPPPP 626
           :|  :: :   : | ||:|| | |::||:|  :       |||  |  :::::  :|  :::
Db    1798 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1847

Qy     627 LTPPLPPSGPSPELARPSSPEHLPRTVPFYPPPRMPPPPPPMPPAPLP PAPPPPPPPV 676
           : :||| |     ||    : ||: :: | |:  |  |::          :: |
Db    1848 VPPPPP PPPHPPLPIAIPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1901

Qy     677 DYPPINPPPTVAAPVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 726
           |    ::: | | :  :  :  :  :   :  : ;:|:
Db    1902 DEAPEVESSTPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1947


RESULT 5
MYSH_ACACA
ID   MYSH_ACACA     STANDARD;      PRT;  1977 AA.
AC   P47808;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   High molecular weight type I myosin I (HMWMI).
OS   Acanthamoeba castellanii (Amoeba).
OC   Eukaryota; Acanthamoebidae; Acanthamoeba.
OX   NCBI_TaxID=5755;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91056121; PubMed=2243110;
RA   Horowitz J.A., Hammer J.A. III;
RT   "A new Acanthamoeba myosin heavy chain. Cloning of the gene and
RT   immunological identification of the polypeptide.";
RL   J. Biol. Chem. 265:20646-20652(1990).
CC   -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60954; AAA27784.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00063; myosin_head; 2.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF03784; MyTH4; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000066; SH3; 1.
DR   ProDom; PD000365; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Actin binding; ATP-binding; Methylation; Multigene family;
KW   SH3 domain.
FT   DOMAIN        1    754       MYOSIN HEAD-LIKE.
FT   DOMAIN      755    782       IQ.
FT   DOMAIN     1519   1577       SH3.
FT   NP_BIND     168    175       ATP (POTENTIAL).
FT   DOMAIN      628    650       ACTIN BINDING (POTENTIAL).
FT   DOMAIN      730    744       ACTIN-BINDING (POTENTIAL).
FT   DOMAIN      277    280       POLY-GLY.
FT   DOMAIN     1471   1476       POLY-ALA.
FT   DOMAIN     1492   1499       POLY-GLN.
FT   DOMAIN     1503   1506       POLY-GLN.
FT   DOMAIN     1506   1511       POLY-GLN.
FT   MOD_RES     119    119       METHYLATION (TRI-) (POTENTIAL).
SQ   SEQUENCE   1577 AA,  177520 MW,  BAF2B8BBAPP7438C CRC64;

  Query Match             4.7%;  Score 192;  DB 1;  Length 1577;
  Best Local Similarity   29.3%;  Pred. No. 0.00028;
  Matches   61;  Conservative   30;  Mismatches   87;  Indels   30;  Gaps    7;

Qy     459 LEPAAEYFEESRQALELQITPDGAAGKDTDSLVQYTKAPIQESLLSLSDPVSKLAVASFLA 517
           ::|||:      :|| ||               : ::||:||  || :|      |||  |
Db    1241 EDPAKAKLLGPEALPIS -  -  ------NPEWTKSPIPTSLTTLDQLAVKSATPLFKN 1287

Qy     518 LMREPMGLQPPSPRFLEMFIVPLYELLKLT-QQEPLFDSIYGQVIKQVTGHPPPEHCTPGWSF 576
           :: ||||:  |       |  :|||: |    :||:|:|||:||||:| :| |:  |:||
Db    1288 VLGPMGDEPLFYPN-  ALAQDLLEQCLAAPELPNEVYCQIIKQLTENPSPQSVTKGWQL 1344

Qy     577 LSLLTGFPPPSTRLMPYLTKFLQDSGPSQFLARSSQEHLQRTVKYGGRRMPPPPGEMKAF 636
           :  |||||      |  ||  |   :      | | | :|| :|   | |
Db    1345 MRCCLQTFFTGEEFANCLEMFLAAFGKDDFYI-  -EMLHDT-QYGDKPTSAPNVEQILA 1399

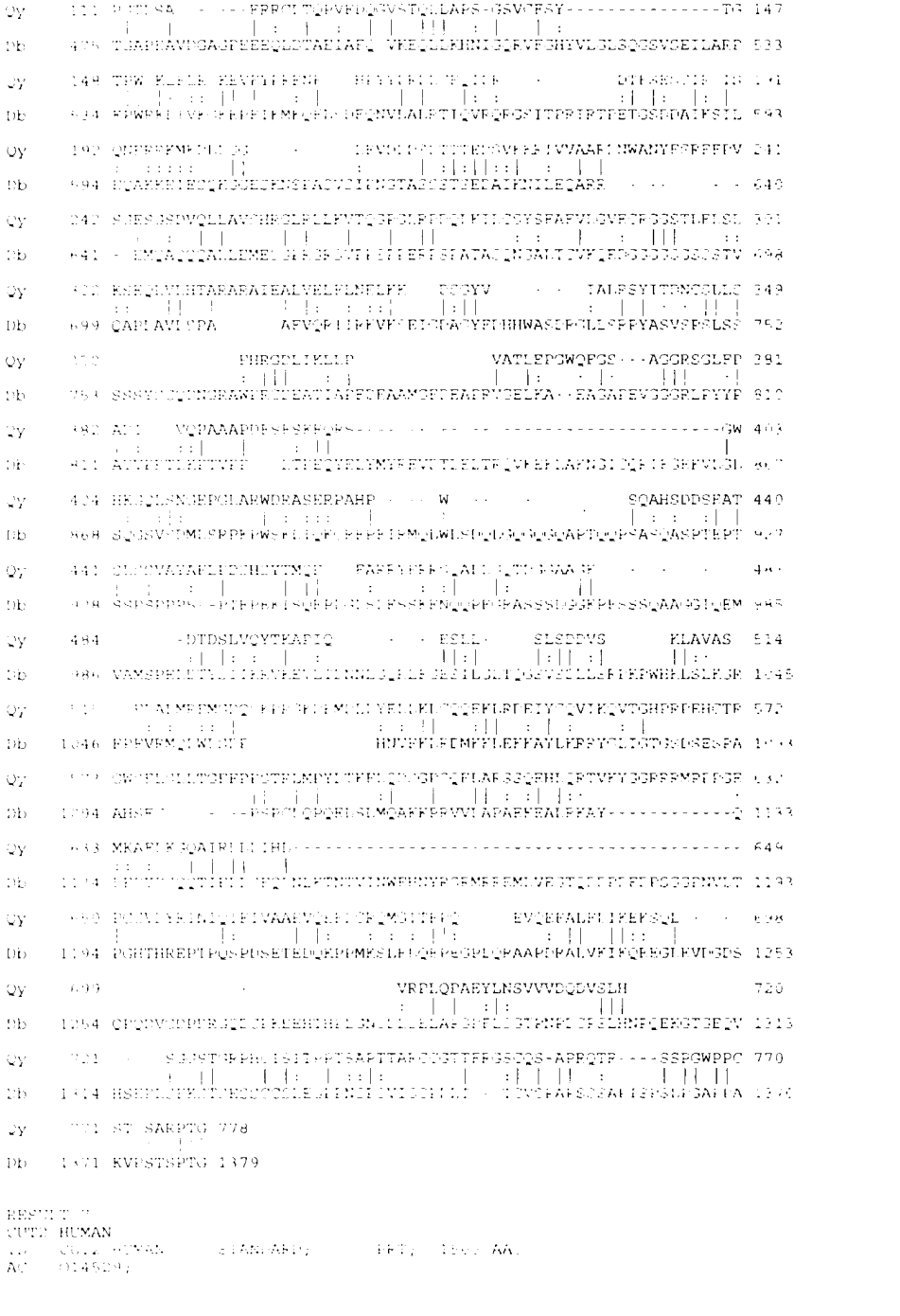
Qy     637 LKQQAIPLLIHIPGGVDYETNTQTFTV 664
           |  |:       ::  | : | |
Db    1400 AKQYVTRI-------DLNQSTQVDTSVV 1420
```

```
RESULT 6
CUT2_MOUSE
ID   CUT2_MOUSE     STANDARD;      PRT:  1426 AA.
AC   P70298;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Homeobox protein Cux-2 (Cut like 2).
GN   CUTL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96394327; PubMed=8798433;
RA   Quaggin S.E., Vanden Heuvel G.B., Golden K., Bodmer R., Igarashi P.;
RT   "Primary structure, neural-specific expression, and chromosomal
RT   localization of Cux-2, a second murine homeobox gene related to
RT   Drosophila cut.",
RL   J. Biol. Chem. 271:22624-22634(1996).
CC   -!- FUNCTION: MAY BE A TRANSCRIPTION FACTOR INVOLVED IN NEURAL
CC       SPECIFICATION. BINDS TO DNA IN A SEQUENCE-SPECIFIC MANNER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO NEURAL TISSUES. EXPRESSED
CC       EXCLUSIVELY IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.
CC   -!- SIMILARITY: CONTAINS 3 CUT DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE CUT HOMEOBOX FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U45665; AAC52762.1; -.
DR   HSSP; P10037; 1AU7.
DR   TRANSFAC; T04456; -.
DR   MGD; MGI:107321; Cutl2.
DR   InterPro; IPR003350; Hmoeo_CUT.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF02376; CUT; 3.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW   Repeat; Coiled coil.
FT   DOMAIN      131    311       COILED COIL (POTENTIAL).
FT   DNA_BIND    483    569       CUT 1.
FT   DOMAIN      587    655       COILED COIL (POTENTIAL).
FT   DNA_BIND    828    917       CUT 2.
FT   DNA_BIND    983   1070       CUT 3.
FT   DNA_BIND   1113   1172       HOMEOBOX.
SQ   SEQUENCE   1426 AA;  154702 MW;  4414D5D0356B984E CRC64;

  Query Match             7.7%;  Score 154.5;  DB 1;  Length 1426;
  Best Local Similarity   18.6%;  Pred. No. 0.056;
  Matches  203;  Conservative  119;  Mismatches  362;  Indels  405;  Gaps   45;

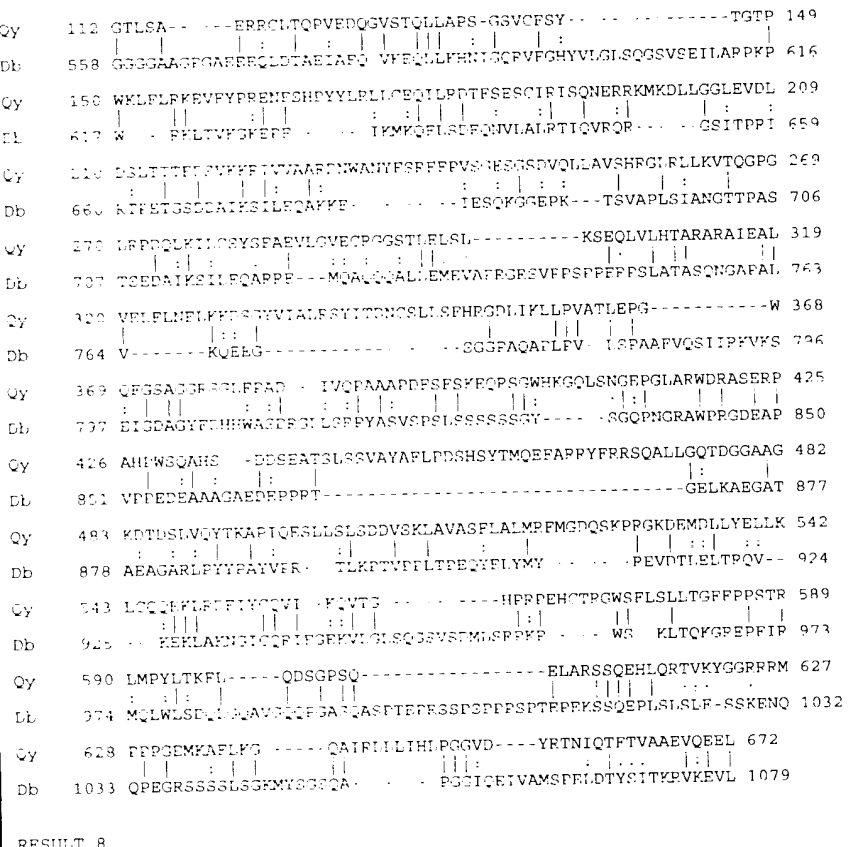
Qy      27 DPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAPPPKAPLQLGPS---------- -- 73
           || ::   |  :      || | | |  ||    ::| |  || ||||
Db     359 DPSEDDSIKGSLGTEPPYPPQLPPPP- GPEDPLSPSIAQFL-LGPSLGPDGPPNPGLSP 415

Qy      74 ---- -------- --- - -- SSIKEKQGPLLDLPGQKLPIAHTPPPPPAPPLPLPED 110
                                  :: | : | || |          || ||  :| ||
Db     416 PPSLAPGEPLAGDSLLSKHMMGPAAPKGETGNLL-AFPPTPYGGAKPPSAPAASVPCPEP 474
```



RESULT 7
CUT2_HUMAN
ID CUT2_HUMAN STANDARD; PRT; 1505 AA.
AC Q14529;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein Cux-2 (Cut like 2) (Fragment).
GN CUTL2 OR KIAA0293.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Nomura N.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 212-1505 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97232008; PubMed=9179496;
RA Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M.,
RA Seki N., Nomura N.;
RT "Construction and characterization of human brain cDNA libraries
RT suitable for analysis of cDNA clones encoding relatively large
RT proteins.";
RL DNA Res. 4:53-59(1997).
CC -!- FUNCTION: MAY BE A TRANSCRIPTION FACTOR INVOLVED IN NEURAL
CC SPECIFICATION. BIND TO DNA IN A SEQUENCE SPECIFIC MANNER (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: CONTAINS 3 CUT DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE CUT HOMEOBOX FAMILY.
CC ---
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AB007831; BAA21962.2; -.
DR TRANSFAC; T04455; -.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR Pfam; PF02376; CUT; 3.
DR ProDom; PD000018; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS50071; HOMEOBOX_2; 1.
KW Transcription; Transcription regulation; Homeobox; DNA binding; Nuclear protein;
KW Repeat; Coiled coil.
FT NON_TER 1 1
FT DOMAIN 213 393 COILED COIL (POTENTIAL).
FT DNA_BIND 563 650 CUT 1.
FT DOMAIN 668 736 COILED COIL (POTENTIAL).
FT DNA_BIND 906 993 CUT 2.
FT DNA_BIND 1057 1144 CUT 3.
FT DNA_BIND 116 176 HOMEO_1.
SQ SEQUENCE 1505 AA; 163792 MW; F6BF9274BCB4BF41E 'B'64;

Query Match 3.7%; Score 153.5; DB 1; Length 1505;
Best Local Similarity 20.5%; Pred. No. 0.07;
Matches 159; Conservative 84; Mismatches 267; Indels 265; Gaps 7;



```
RESULT 8
MY10_HUMAN
ID   MY10_HUMAN     STANDARD;      PRT;  2058 AA
AC   Q9HD67; Q8NYM3; Q9P110; Q9P111; Q9UHP6;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin X.
GN   MYO10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20442246; PubMed=10984435;
RA   Berg J.S., Derfler B.H., Pennisi C.M., Corey D.P., Cheney R.E.;
RT   "Myosin-X, a novel myosin with pleckstrin homology domains, associates
RT   with regions of dynamic actin.";
RL   J. Cell Sci. 113:3439-3451(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rogers M.S., Strehler E.E.;
RT   "Identification of myosin X as a specific binding partner for the
RT   tumor sensative calmodulin-like protein.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colorectal adenocarcinoma;
RA   Takada T., O'Farrell T.J., Anderson J.T., Pourmotabbed T.;
RT   "Cloning of human myosin X.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 347-495 FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=20079149; PubMed=10610710;
RA   Rojas K., Serrano de la Pena L., Gallardo T., Simmons A., Nyce K.,
RA   McGrath P., Considine E., Vasko A.J., Peterson E., Grady D., Cox E.,
RA   Andrew L.J., Lovett M., Overhauser J., Williams C.J.;
RT   "Physical map and characterization of transcripts in the candidate
RT   interval for familial chondrocalcinosis at chromosome 5p15.1.";
RL   Genomics 62:177-183(1999).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC       SIMILARITY). PLAYS A ROLE IN REGIONS OF DYNAMIC ACTIN.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF247457; AAF68025.2; -.
DR   EMBL; AF234532; AAF37875.1; -.
DR   EMBL; AF132021; AAF36524.1; -.
DR   EMBL; AF132020; AAF36525.1; -.
DR   EMBL; AF184153; AAF17363.1; -.
DR   HSSP; P08799; 1MND.
DR   GeneW, HGNC:7593; MY010.
DR   MIM; 601481; -.
DR   InterPro; IPR000299; Band_4.1
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 2.
DR   Pfam; PF00169; PH; 3.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00784; MyTH4; 2
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 1.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 1
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS00660; BAND_41_1, FALSE_NEG.
DR   PROSITE; PS00661; BAND_41_2; FALSE_NEG.
DR   PROSITE; PS50057; BAND_41_3; 1.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Myosin; ATP-binding; Actin-binding, Coiled coil; Repeat
FT   DOMAIN        1    727       HEAD OR MOTOR DOMAIN.
FT   DOMAIN      742    763       IQ 1.
FT   DOMAIN      764    787       IQ 2.
FT   DOMAIN      788    817       IQ 3.
FT   DOMAIN     1212   1310       PH 1.
FT   DOMAIN     1392   1497       PH 2.
FT   DOMAIN     1697   1981       BAND 4.1-LIKE.
FT   DOMAIN      729    943       COILED COIL (POTENTIAL).
FT   CONFLICT     98     98       S -> P (IN REF. 3).
FT   CONFLICT    148    148       H -> Y (IN REF. 2).
FT   CONFLICT    256    256       G -> W (IN REF. 1).
```

```
FT   CONFLICT     424    424       W -> R (IN REF. 2).
FT   CONFLICT    1186   1186       G -> C (IN REF. 3).
FT   CONFLICT    1663   1663       T -> S (IN REF. 3).
SQ   SEQUENCE   2058 AA;  237389 MW;  B4E19BF92D87F3AC CRC64;

  Query Match              3.7%;  Score 152.5;  DB 1;  Length 2058;
  Best Local Similarity   24.3%;  Pred. No. 0.11;
  Matches   56;  Conservative   42;  Mismatches   89;  Indels   37;  Gaps    7;

Qy     488 LVQYTKAPIQESLL-  -- - ---  -- --  SLSDDVSKLAVACPLALMPPMGVQQKPP 539
           ::::|  |:   ||                  :| |:  |:          |   |
Db    1532 ILRHTHHPLHSPLLPLPYGLINLNLIKFMGYIIQDPAIKI        - ENSLQQLES 1581

Qy     540 GKIEMDLLYELLKLCQG-EFIRDFIYCQVIKQVTSHEK     S-- TPWWPLCLLTPP 596
           | : ::  :|:         ||||:|||:|||    | |       |:  |  |: |: |
Db    1582 MSIPIFIIQTIL-QHCHSLRPLRIELYIQIFLDPFVHH 33% DNIYSWLILTTLSCTFIL 1641

Qy     597 PPSTPIMPYLTPF    LQPQPSPLAPSQPHLQWIVKAGAAAAMPPPPMKAPLPQQ 649
           ||  :: || ||     ::: | |: : :  :  | |  :  |  |::|  : |
Db    1642 LPSPIIPYFPFPLPVSIPFPIPVPFPVSIPFPIPVPFPVSIPFPIPAVPAQ 1695

Qy     650 AIPLLLIHLPGSVNYFTNIQTFTVAAFVQEELCKQMGITEPQEVLEPALF 699
           :  :|  ||   :  : | | | |:| | : :  :  :  |||| |
Db    1696 EM TSTVFHGGGSCKITINSHTTAGEVTEKLIRGLAMEDSRNM  FALF 1739


RESULT 9
MYI0_BOVIN
ID   MYI0_BOVIN     STANDARD;      PRT;  2052 AA.
AC   P79114;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin X.
GN   MYO10.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=20412240; PubMed=10994425;
RA   Berg J.S., Powler B.H., Pennisi C.M., Corey D.P., Cheney R.E.;
RT   "Myosin-X, a novel myosin with pleckstrin homology domains, associates
RT   with regions of dynamic actin.";
RL   J. Cell Sci. 113:3439-3451(2000).
CC   -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS HAVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PROPOSED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC       SIMILARITY). PLAYS A ROLE IN REGIONS OF DYNAMIC ACTIN.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE MOTOR/ATP HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/Announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U55042; AAB39486.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000299; Band_4.1
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00169; PH; 3.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00784; MyTH4; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   Prelom; PS50333; myosin_head; 1.
DR   SMART; SM00295; B41; 1.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 1.
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS50057; BAND_4_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_4_1_2; FALSE_NEG.
DR   PROSITE; PS00657; BAND_4_1_3; 1.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.
FT   DOMAIN       1    727       HEAD OR MOTOR DOMAIN.
FT   DOMAIN     742    763       IQ 1.
FT   DOMAIN     764    787       IQ 2.
FT   DOMAIN     788    817       IQ 3.
FT   DOMAIN    1206   1304       PH 1.
FT   DOMAIN    1386   1491       PH 2.
FT   DOMAIN    1691   1975       BAND 4.1 LIKE.
FT   DOMAIN     800    941       COILED COIL (POTENTIAL).
SQ   SEQUENCE   2052 AA;  235837 MW;  43DF13424B4B2C28 CRC64;

  Query Match              3.7%;  Score 152.5;  DB 1;  Length 2052;
  Best Local Similarity   24.3%;  Pred. No. 0.12;
  Matches   56;  Conservative   42;  Mismatches   89;  Indels   45;  Gaps    9;

Qy     488 LVQYTKAPIQESLL-  -- - ---  -- --  SLSDDVSKLAVACPLALMPPMGVQQKPP 539
           ::::|  |:   ||                  :| |:  |:          |   |
Db    1532 ILRHTHHPLHSPLLPLPYGLINLNLIKFMGYIIQDPAIKI        - ENSLQQLES 1581

Qy     540 GKIEMDLLYELLKLCQG-EFIRDFIYCQVIKQVTSHEK     S-- TPWWPLCLLTPP 596
           | : ::  :|:         ||||:|||:|||    | |       |:  |  |: |: |
Db    1582 MSIPIFIIQTIL-QHCHSLRPLRIELYIQIFLDPFVHH 33% DNIYSWLILTTLSCTFIL 1641

Qy     597 PPSTPIMPYLTPF    LQPQPSPLAPSQPHLQWIVKAGAAAAMPPPPMKAPLPQQ 649
           ||  :: || ||     ::: | |: : :  :  | |  :  |  |::|  : |
Db    1642 LPSPIIPYFPFPLPVSIPFPIPVPFPVSIPFPIPVPFPVSIPFPIPAVPAQ 1695

Qy     650 AIPLLLIHLPGSVNYFTNIQTFTVAAFVQEELCKQMGITEPQEVLEPALF 699
           :  :|  ||   :  : | | | |:| | : :  :  :  |||| |
Db    1696 EM TSTVFHGGGSCKITINSHTTAGEVTEKLIRGLAMEDSRNM  FALF 1739


RESULT 10
WAS1_HUMAN
ID   WAS1_HUMAN     STANDARD;      PRT;   559 AA.
AC   Q92558;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Wiskott-Aldrich syndrome protein family member 1 (WASP family
DE   protein member 1) (Verprolin homology domain-containing protein 1)
GN   WASF1 OR SCAR1 OR KIAA0269.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98402542; PubMed=9732292;
RA   Bear J.E., Rawls J.F., Saxe C.L. III;
RT   "SCAR, a WASP-related protein, isolated as a suppressor of receptor
RT   defects in late Dictyostelium development.";
RL   J. Cell Biol. 142:1325-1335(1998).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99108169; PubMed=9880097;
RA   Machesky L.M., Insall R.H.;
RT   "Scar1 and the related Wiskott-Aldrich syndrome protein, WASP,
RT   regulate the actin cytoskeleton through the Arp2/3 complex.";
RL   Curr. Biol. 8:1347-1356(1998).
CC   -!- FUNCTION: DOWNSTREAM EFFECTOR MOLECULES INVOLVED IN THE
CC       TRANSMISSION OF SIGNALS FROM TYROSINE KINASE RECEPTORS AND SMALL
CC       GTPASES TO THE ACTIN CYTOSKELETON.
CC   -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX.
CC   -!- SUBCELLULAR LOCATION: DOT-LIKE PATTERN IN THE CYTOPLASM
CC       CONCENTRATED IN FAC REGULATED MEMBRANE RUFFLING AREAS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN, LOWLY EXPRESSED IN
CC       TESTIS, OVARY, COLON, KIDNEY, PANCREAS, THYMUS, SMALL INTESTINE,
CC       AND PERIPHERAL BLOOD
CC   -!- DOMAIN: BINDS ARP2/3 COMPLEX THROUGH THE C-TERMINAL REGION AND
CC       ACTIN THROUGH VERPROLIN HOMOLOGY (VPH) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 WH2 DOMAIN.
CC   ---------- . . ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D87459; BAA13399.1; -.
DR   EMBL; AF134305; AAD33052.1; -.
DR   Genew; HGNC:12732; WASF1.
DR   MIM; 605035; -.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1
DR   SMART; SM00246; WH2; 1
KW   Actin-binding.
FT   DOMAIN       278     283       POLY-PRO
FT   DOMAIN       322     332       POLY-PRO
FT   DOMAIN       348     359       POLY-PRO
FT   DOMAIN       369     374       POLY-PRO
FT   DOMAIN       424     435       POLY-PRO
FT   DOMAIN       497     514       WH2.
SQ   SEQUENCE   559 AA;  61652 MW;  44B4527B0B77BC6E CRC64;

  Query Match             3.5%;  Score 143;  DB 1;  Length 559;
  Best Local Similarity   24.5%;  Pred. No. 0.087;
  Matches   67;  Conservative   36;  Mismatches   97;  Indels   74;  Gaps   15;

Qy       1 MYQSRF-----SFVFVPVQPSRPPKAFLPKIDPKDEALAKLGINGAHSS--PPMLSPSPG 53
           : ::|||     |  || | |: ||        |   ||: : : :|  || : | |
Db     303 LIENRPQSPATGRTPVFVSPTPPPPP-----PPLPSALSTSSLRASMTSTPPPPVPPPP- 356

Qy      54 KGPPPAVA   - -PRPKAPLQLGP -- - -  SSSIPEKQGPLLDLPGQKLPIAHT- 96
            |||| |      | | |||: |          .  . :   |:     :  |.  |
Db     357 --PPPATALQAPAVPPPPAPLQIAPGVLHPAPPPIAPPLVQPSPPV----APAAPVQPTV 410

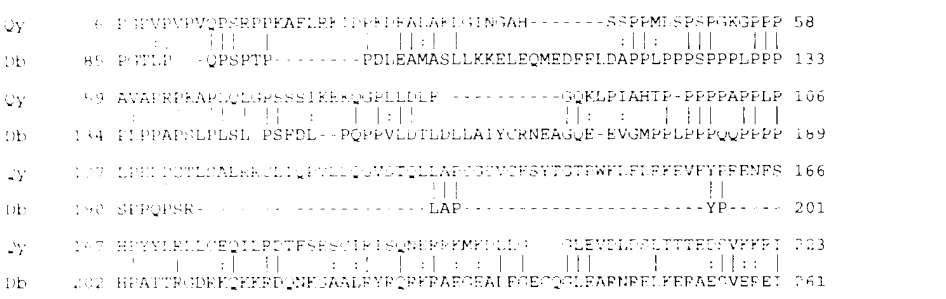
Qy      97 -----     ----PPPPPAPPLP   - LPEDPGTLSAEPPQLTQPVEDQGV-STQLLAPS 138
                       |||||  ||||     |  | |..|    |  |    |- |    ||
Db     411 PVHPLPQGEVQGLPPPPPPPPLPPPGIRPSSPVTVTA----LAHP--PSGLHPTPSTAPG 464

Qy     139 GSVGPSYTGTPWKLFLPVFVFYPPENPSHPYYLRILGE -QILRDTPSESCIRISQNERR 196
           |  |  |      :|    |  ||  | :: .   :| :  :          :.|
Db     465 PHVPLMPPSPP---  -SQVIPASEPKEHPSTLPVISDAPSVLLEAIRKGIQLRKVEEQR 518

Qy     197 KMKDLIGGLEVDLAALTTTEDSVKFPIVVAAPDN 230
           : :    :| |: ::      : :|| |   |:
Db     519 EQEAKHFRIENDVATI   ----LSRRIAVEYSDS 546
```

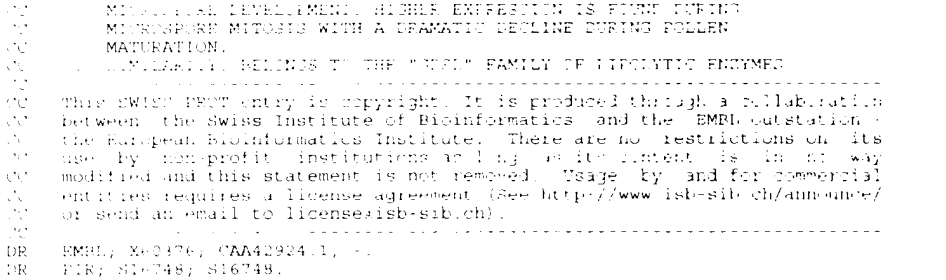
```
RESULT 11
ATF5_HUMAN
ID   ATF5_HUMAN     STANDARD;      PRT;   282 AA.
AC   Q9Y2D1; Q9UNQ3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cyclic-AMP-dependent transcription factor ATF-5 (Activating
DE   transcription factor 5) (Transcription factor ATFx).
GN   ATF5 OR ATFX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A
RX   MEDLINE=20559615; PubMed=11087924;
RA   White J.H., McIlhinney R.A.J., Wise A., Ciruela F., Chan W.Y.,
RA   Emson P.C., Billinton A., Marshall F.H.;
RT   "The GABAB receptor interacts directly with the related transcription
RT   factors CREB2 and ATFx.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:13967-13972(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kohroki J., Tanaka K.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 161-282 FROM N.A.
RX   MEDLINE=99303793; PubMed=10373550;
RA   Pati D., Meistrich M.L., Plon S.E.;
RT   "Human Cdc34 and Rad6B ubiquitin-conjugating enzymes target repressors
RT   of cyclic AMP-induced transcription for proteolysis.";
RL   Mol. Cell. Biol. 19:5001-5013(1999).
CC   -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
CC       (CONSENSUS: 5'GTGACGT(A/C)(A/G)-3'), A SEQUENCE PRESENT IN MANY
CC       VIRAL AND CELLULAR PROMOTERS.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER (POTENTIAL). INTERACTS WITH GABAB
CC       RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
CC   ---------- . . ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF305687; AAG22558.1; -
DR   EMBL; AB021663; BAA78477.2; -
DR   EMBL; AF101388; AAD28370.1; -.
DR   TRANSFAC; T04877; -.
DR   Genew; HGNC:790; ATF5.
DR   MIM; 606398; -.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; DNA binding; Activator; Nuclear protein;
KW   Multigene family.
FT   DOMAIN       123     139       POLY-PRO.
FT   DOMAIN       186     194       POLY-PRO.
FT   DNA_BIND     210     230       BASIC MOTIF.
FT   DOMAIN       236     250       LEUCINE-ZIPPER (PROBABLE).
FT   CONFLICT     161     163       LLA -> RHE (IN REF. 3).
SQ   SEQUENCE   282 AA;  30674 MW;  DDB2F9977AC21FA0 CRC64;

  Query Match             3.4%;  Score 140.5;  DB 1;  Length 282;
  Best Local Similarity   27.1%;  Pred. No. 0.05;
  Matches   65;  Conservative   18;  Mismatches   72;  Indels   57;  Gaps   12;
```



RESULT 12
APG_BRANA
ID APG_BRANA STANDARD; PRT; 449 AA.
AC P46603;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Anther-specific proline rich protein APG (Protein CEX) (Fragment).
GN APG OR CEX.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94034990; PubMed=8401599;
RA Edlund A.F., Potter S.D., Blundell P.P., Robinson C.W., Kumar A.,
RA Draper J., Scott R.;
RT "Gametophytic and sporophytic expression of an anther specific
RT Arabidopsis thaliana gene.";
RL Plant J. 3:111-120(1993).
CC -!- TISSUE SPECIFICITY: FOUND IN ANTHER, ONLY IN MALE FERTILE PLANTS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC MATURATION.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X60376; CAA42924.1; -.
DR PIR; S16748; S16748.
DR InterPro; IPR001087; Lipase_GDSL.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR PROSITE; PS01098; LIPASE_GDSL_SER; 1.
FT NON_TER 1 1
FT ACT_SITE 132 132 BY SIMILARITY.
FT ACT_SITE 428 428 POTENTIAL.
SQ SEQUENCE 449 AA; 48779 MW; 9E8F0A3A8J98EA15 CRC64;

Query Match 3.38%; Score 138; DB 1; Length 449;
Best Local Similarity 22.7%; Pred. No. 2.1e-13;
Matches 93; Conservative 26; Mismatches 109; Indels 147; Gaps 20;



RESULT 13
FMN1_MOUSE
ID FMN1_MOUSE STANDARD; PRT; 1468 AA.
AC Q38940;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Formin 1 isoforms I/II/III (Limb deformity protein).
GN FMN OR LD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney, and Testis;
RX MEDLINE=97267291; PubMed=9192150;
RA Wychik B.F., Maas R.L., Zeller R., Vogt T.F., Leder P.;
RT "'Formin' proteins deduced from the alternative transcripts of the
RT limb deformity gene.";
RL Nature 387:832-835(1997).
RN [2]
RP ALTERNATIVE SPLICING.
RX MEDLINE=98224459; PubMed=9119367;
RA Wang C.C., Chan D.C., Leder P.;
RT "The mouse formin (Fmn) gene: genomic structure, novel exons, and
RT genetic mapping.";
RL Genomics 39:303-311(1997).
RN [3]
RP PHOSPHORYLATION.
RX MEDLINE=93296176; PubMed=8516300;
RA Vogt T.F., Jackson-Grusby L., Puch J., Leder P.;
RT "Formins: phosphoprotein isoforms encoded by the mouse limb deformity
RT locus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5554-5558(1993).
CC -!- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
CC AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
CC INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1A (SHOWN HERE), 1B,
CC II, III AND IV (AC Q05859); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC A VARIATION IN SPLICING IN FMN AMONG DIFFERENT TISSUES AND
CC DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
CC -!- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
CC LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
CC GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
CC -!- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
CC IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL

```
CC       COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
CC       KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
CC       THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
CC   -!- PTM: PHOSPHORYLATED ON SERINE AND POSSIBLY THREONINE RESIDUES.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FH1) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. CAPPUCCINO
CC       SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ----------------------------------------------------------------------
DR   EMBL; X53599; CAA37668.1; -.
DR   PIR; S11515; S11515.
DR   MGD; MGI:101815; Fmn.
DR   InterPro; IPR003104; FH2.
DR   InterPro; IPR001265; Formin.
DR   Pfam; PF02181; FH2; 1.
DR   PRINTS; PR00928; FORMIN
DR   SMART; SM00498; FH2; 1.
KW   Nuclear protein; Developmental protein; Alternative splicing;
KW   Phosphorylation; Coiled coil
FT   DOMAIN      777    [illegible]  COILED COIL (POTENTIAL)
FT   DOMAIN      870    970          FH1 (PRO-RICH).
FT   DOMAIN      985    1404         FH2
FT   DOMAIN     1305    1378         COILED COIL (POTENTIAL).
FT   DOMAIN      198     203         POLY-SER.
FT   DOMAIN      861     864         POLY-SER.
FT   DOMAIN      885     892         POLY-PRO.
FT   DOMAIN      911     925         POLY-PRO.
FT   DOMAIN      929     940         POLY-PRO.
FT   DOMAIN      951     960         POLY-PRO.
FT   DOMAIN      968     978         POLY-SER.
FT   DOMAIN      977     981         MISSING (IN ISOFORM IB)
FT   VARSPLIC   1252    1287         MISSING (IN ISOFORM II).
FT   VARSPLIC    625     722         IA -> SV (IN ISOFORM III).
FT   VARSPLIC    626     627         MISSING (IN ISOFORM III).
FT   VARSPLIC    628    1468         MISSING (IN ISOFORM III).
SQ   SEQUENCE   1468 AA;  161899 MW;  EF2FB1E9CA9DAF43 CRC64;

  Query Match              3.2%;  Score 134;  DB 1;  Length 1468;
  Best Local Similarity   20.9%;  Pred. No. 1.1;
  Matches  105;  Conservative   51;  Mismatches  138;  Indels  210;  Gaps   25;

Qy        6 PGPVPVPVQPSRPPPAPLPYIDPKPDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAP-RP 64
            | |:| |: |  ||                              || | | |   | | | |  |
Db      876 PPPLPPPLIPPPPP----------------------LPPGLGPLPPAPPIPPVCPVSP 911

Qy       65 KAPLQLGPSSSIKEKIGPILPLPQQKLPIAHTPPPPPAPPL----------------PL 107
            |    | : :    ||                 |||||  |||                  |
Db      912 PPPPPPPPPTPVPPSDGP-------------PPPPPPPPPLPNVLALPNSGGPPPPPPPP 958

Qy      108 PELPGTLSAEERQLLPVEPQGVSTQLLAPSGSVCFSYTGTPWKLPLPKEVPYPPENFSH 167
            |  ||           |    |:|  | : |                          |||.
Db      959 PPPPG---------LAPPPPGLSFGLSSSSSQ------------ .. . YPRKPAIE 990

Qy      168 P-------YYLRL------- --------LCEQILRDT------FSESCIR--------- 189
            |       |: |:                | |  :|||      ||:   :
Db      991 PSCPMKPLYWTRIQINDKSQDAAPTLWDSLEEPHIRDTSEFEYLFSKDTTQQKKKPLSEA 1050

Qy      190 -ISQNERPKMKDLLGG-----------LEVDL----DSLTTTEDSVKKRIVVA------ 226
             :|: :|:  || |              | :::     :: | :|||      :|
Db     1051 YEKPCNKVKKIIKILDGKPSQTVGILISSLHLEMKDIQQAIFTVDDSVVDLETLAALYENR 1110

Qy      227 ARDNWANYFSPPPPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEV 286
            |:::          |  |   |: ||    .  |   :: | |       ||    |:|
Db     1111 AQELELTKIRKYYEISRE- EELFLIDPEEQFIH -ELAQIINFAEPAQCIIFFAVFSE- 1165

Qy      287 LGVECRGGST-----LELSLKSEQLVLHTARARAIEALVELFLNEL------------ 327
                 | |      .|:  .: : :||     : | ||:  | | :
Db     1166  - ----GITSLHRKVEIVTRASKGLLHMKSVKDILALILAFSNYMN-SNFTFQQAPTYS 1218

Qy      328 ------- ---FFDSG-----YVIA--LRSYITDNCSLLSFH-      RG-DLIFILPVA 362
                     .|:|      ||:   || | | |      |       || || |
Db     1219 ISILFELFDVKSPNGGMNLVDYVVKYYLEYY--DQC-  KHHLQEAS-PFSKDLFSLYFHI 1273

Qy      363 TLEPGWQPG-----SAGGRSGLFP 381
              : |  : |      ||    :||
Db     1274 AVHPQPKSGLELKQEAGTDKSVFP 1297


RESULT 14
PXRB_HUMAN
ID   PXRB_HUMAN     STANDARD;      PRT:   533 AA.
AC   P28702; P28703;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Retinoic acid receptor RXR-beta.
GN   PXRB OR NR2B2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=92253286; PubMed=1315058;
RA   Fleischhauer K., Park J.H., Disanto J.F., Marks M.S., Gra[illegible] K.,
RA   Yang S.Y.;
RT   "Isolation of a full-length cDNA clone encoding a N terminally
RT   variant form of the human retinoid X receptor beta.";
RL   Nucleic Acids Res. 20:1801-1801(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=92127595; PubMed=1310259;
RA   Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,
RA   Zacharewsi T., Chen J.Y., Staub A., Garnier J.M., Mader S.,
RA   Chambon P.;
RT   "Purification, cloning, and RXR identity of the HeLa cell factor with
RT   which RAR or TR heterodimerizes to bind target sequences
RT   efficiently.";
RL   Cell 68:377-395(1992).
RN   [3]
RP   REVISIONS.
RX   MEDLINE=93046692; PubMed=1330328;
RA   Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,
RA   Zacharewsi T., Chen J.Y., Staub A., Garnier J.M., Mader S.,
RA   Chambon P.;
RL   Cell 71:887-887(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Numasawa T., Koga H., Ueyama K., Maeda S., Sakou T., Harata S.,
RA   Leppert M., Inoue I.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Corella A., Vergara A., Paez G., de Miguel C., Prieto I.;
RT   "Molecular cloning and characterization of the human HRXRB gene and 5'
RT   flanking region.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Tubby B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
```

```
CC       9-CIS RETINOIC ACID (9C-RA).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; a long form (shown
CC       here) and a short form, arose by both alternative splicing
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C TERMINAL STEROID BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR1 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63629; CAA45087.1; -.
DR   EMBL; M84820, AAA60233.1, -.
DR   EMBL; AF067936, AAC18599.1, -.
DR   EMBL; AF120161, AAD13794.1, -.
DR   EMBL; AL031228; CAA20239.1; -.
DR   EMBL; BC001167; AAH01167.1; -.
DR   PIR; S22647; S22647.
DR   PIR; S39753; S39753.
DR   HSSP; P19793; 2NLL.
DR   TRANSFAC; T01334; -.
DR   Genew; HGNC:10478; RXRB.
DR   MIM; 180245; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Steroid_recep_fam.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA binding; Nuclear protein;
KW   Zinc finger; Multigene family; Alternative splicing.
FT   DOMAIN        1    504       MODULATING (BY SIMILARITY).
FT   DNA_BIND     505    570       NUCLEAR RECEPTOR-TYPE
FT   ZN_FING      205    225       C4-TYPE.
FT   ZN_FING      241    265       C4-TYPE.
FT   DOMAIN       271    338       HINGE
FT   DOMAIN       339    503       LIGAND-BINDING (BY SIMILARITY).
FT   DOMAIN        61    198       PRO RICH.
FT   CONFLICT     112    112       S -> T (IN REF. 2).
SQ   SEQUENCE   533 AA;  59721 MW;  DCDA4FE8A7C9A4A4 CRC64;

  Query Match             3.2%;  Score 133;  DB 1;  Length 533;
  Best Local Similarity   22.6%;  Pred. No. 0.35;
  Matches  93;  Conservative   46;  Mismatches  153;  Indels  109;  Gaps   19;

Qy        4 SPPLPVPPTIVQFSPPPKAPLRKIPPTPAIAPLGINGAHGSPPMLGP------------ 50
            :|  | |  :|   |    |  |   || :||   |||  |
Db       86 SSPNPLPQSVPPPSPPGPPL----PPSTA-PSLGGGGAPPPPPMPPPPLGGPPPVISGGM 140

Qy       51 SPGKAPP  -----PAVAPPPKAPLQL-----GPSSSIPEPLQPLLDLPGQFLPIAHTP 97
            :: |     |  :|  : : |     ||   :|    :|: : |      | |
Db      141 GSPGLPPPAPPGPFSGPVSSPQINSTVSLPQGGSGPPELVPP---PVLGVRG-- --LHCP 193

Qy       98 PPPPAPPLPVFPPPQTLPAEPKTLPVTLPVTLLAPGGVTPYTTTPWPLFLPKE 151
            ||  |          : :| |       |     :  |  |      :|  :||:
Db      194 PPPGGGG-   --      AGPRLPAIPPPPPGPPHYGVYSPPGPKVF      PPPTIPPP 236

Qy      152 VPY  -          --PPENPSHPYYLPLLQEQIPFTPSPGTIPIQLNFPPPMPL--- 201
            :|           :  | : |   : |:           | ||:: ||
Db      237 PPP PPPPIMPPPTVVPPPLPMPPLYYPYPPLAPGMPPPAV  -    - QPPPQMPGPGKPMPP 283

Qy      202    LKDL  PVTLLQLTTPEQVPEPTTVTAPLNWAQTPSRPFPVEGEGG         - 247
```

```
                ||   |: :| :   | :|:::                    :|  :|
Db      289 GEGAGGAPEEMPVDRILEAELAVEQKSDQGVEGPGS          TGGPGDSLNLPVTN 344

Qy      248    --TVQLLAVSPPPNPLIPVTQGPNIPTP_IPILPVYTPAPVI        -DVECPKNT 296
               | ||  :      |:   : |    |  ||    :  :|          :: . |
Db      345 GLAALEQLETLVEWARRIIHPEGGLP   LDQQVILPAGWNLLIAPLWRQSIPVPDHLF 404

Qy      297 LELSLPAE_LVQPTAPAPAI  EALVLIP  LPPIPP   TPTPALI PITPN 344
            |    |   :|:|   ||    | ||    : ::: |   :  :  :  |
Db      405 LAPGRHVPPPSAHLAJVGATSLPVLPLVPMPPMPMQPTPPLPPPPALLIPN 448


RESULT 15
FM14_MOUSE
ID   FM14_MOUSE     STANDARD;      PRT;  1206 AA.
AC   Q05859;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Formin 1 isoform IV (Limb deformity protein).
GN   FMN OR LD.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=92112033; PubMed=1439380;
RA   Grusby-Jackson L., Kuo A., Leder P.;
RT   "A variant limb deformity transcript expressed in the embryonic mouse
RT   limb defines a novel formin.";
RL   Genes Dev. 6:29-37(1992).
CC   -!- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB AND MAY HAVE A
CC       FUNCTION IN DIFFERENTIATED CELLS OF BE INVOLVED IN MAINTAINING
CC       SPECIFIC DIFFERENTIATED STATES.
CC   -!- ALTERNATIVE PRODUCTS: At least 4 isoforms; A (AC Q05860), IB (AC
CC       Q05860), II (AC Q05860) AND IV (THIS ENTRY) ARE
CC       PRODUCED BY ALTERNATIVE SPLICING. A VARIATION IN SPLICING IS SEEN
CC       AMONG DIFFERENT TISSUES AND DIFFERENT SIZE TRANSCRIPTS EXIST
CC       WITHIN ANY ONE TISSUE.
CC   -!- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT the EMBRYO BUT
CC       HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.
CC    !  DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL
CC       ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING
CC       LIMB BUD.
CC   -!- PTM: PHOSPHORYLATED ON SERINE AND POSSIBLY THREONINE RESIDUES.
CC    !  SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FH1) DOMAIN.
CC    !  SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. DAPPLE/DIAPH
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62379; CAA44244.1; -.
DR   PIR; S24407; S24407.
DR   MGD; MGI:101815; Fmn.
DR   InterPro; IPR003104; FH2.
DR   InterPro; IPR002200; Formin.
DR   Pfam; PF02181; FH2; 1.
DR   PRINTS; PR00928; FORMIN.
DR   SMART; SM00498; FH2; 1.
KW   Nuclear protein; Developmental protein; Alternative splicing;
KW   Phosphorylation; Coiled coil.
FT   DOMAIN      419    443       COILED COIL (POTENTIAL).
FT   DOMAIN      497    566       COILED COIL (POTENTIAL).
FT   DOMAIN      644    744       FH1 (PRO-RICH).
```

```
FT   DOMAIN      759   1164        FH2.
FT   DOMAIN     1043   1110        COILED COIL (POTENTIAL).
FT   DOMAIN      635    638        POLY-SER.
FT   DOMAIN      751    755        POLY-SER.
SQ   SEQUENCE   1206 AA;  133464 MW;  4DFB38CB52BD8EE7 CRC64;

  Query Match              3.2%;  Score 132;  DB 1;  Length 1206;
  Best Local Similarity   20.3%;  Pred. No. 1.2;
  Matches   87;  Conservative   49;  Mismatches  125;  Indels  168;  Gaps   18;

Qy       6 PGPVPVPVQPSRPPKAPLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAP-RP 64
           | |:| |: |  ||                                          | | | | | | | | |
Db     650 PPPLPPPLIPPPPP----------------------LPPGLGPLPPAPPIPPVCPVSP 685

Qy      65 KAPLQLGPSSSIKPKQGPLLDLPGQKLPIAHTPPPPPAPPL----------------PL 107
             |    | : :    ||              |||||  |||                  |
Db     686 PPPPPPPPPTPVPPSDGP----------PPPPPPPPLPNVLALPNSGGPPPPPPPP 732

Qy     108 PEDPGTLSAFPPDLTQLVELQGVSTQLLAPSGSVCPSYTGTPWFLPLPKEVPYPPENFSH 167
           |  ||           |    |:|  | : |                    |||:
Db     733 PPPPG---------LAPPPPPGLSPGLSSSSSQ----------------YPRKPAIE 764

Qy     168 P------ YYLRL     --------LCEQILRDT------FSESCIR--------- 189
           |      |: |:                | |  :|||      ||:   :
Db     765 PSCPMKPLYNTRIQINPKSQSAAPTLWPSLEEPHIPDTSEPEYLPSKDTTQQKKPLSEA 824

Qy     190 -ISQNEEPPMKPLIGG     --      LEVDL--  DSLTTTEDSVKKPIVVA------ 226
            .|. :|:  || |              | ...     :: | :|||     :|
Db     825 YEKPMKVPKIIPLLPGKPSQTVGILISSPHLEMKDIQQAIPTVDDSVVDLETLAALYENR 884

Qy     227 ARDNWANYFSRFFPVSGEGSDVQLLAVSHPGLPLLKVTQQPGLPPDQLKILCSYSPAEV 286
           |:..     ...  | |   |::||     . |   :: | |          |:    |:|
Db     885 AQEDELTKIPKYYETSPF-EDLKLIDKPPQFLH-ELAQIPNFAERAQCIIPRAVFSE- 939

Qy     287 LGVECRGGST---  LELSLKSEQLVLHTAPAPAIEALVELPLNELK-------KDSGYV 334
                 | |      :|:  :: : :||    : | ||:  | | :         :  ||
Db     940       SITCLHRKVEIVTPASKGLLHMKSVKDILALILAPGNYMNGGNPTPGQADGYS 992

Qy     335 IALRSYITD 343
           : :   : |
Db     993 LEILPKLKD 1001


Search completed: January 10, 2003, 02:19:53
Job time : 38.7203 secs
```

GenCore version 5.1.2
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 01:06:34 ; Search time 50.9672 Seconds
(without alignments)
3177.594 Million cell updates/sec

Title: US-09-803-126-6
Perfect score: 4129
Sequence: 1 MTVSKIGEVEVEVQESEEEF ... WPPCSTSARPTGIPPQGPTC 786

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206947115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

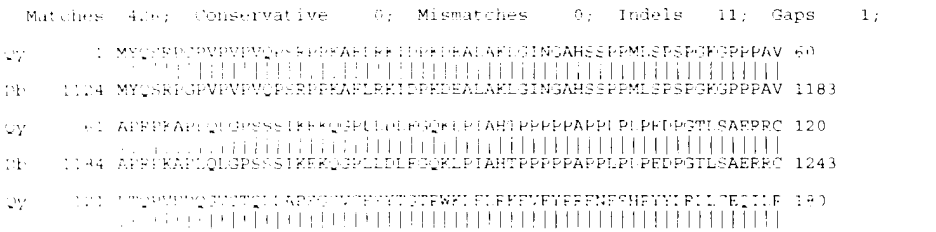
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2217.5	53.7	1560	4	Q96JP2	Q96jp2 homo sapien
2	1346	32.6	487	4	Q9H614	Q9h614 homo sapien
3	1223	29.6	461	4	Q8TCJ6	Q8tcj6 homo sapien
4	805.5	19.5	417	11	Q8R1I5	Q8r1i5 mus musculu
5	545	13.2	2810	5	Q20456	Q20456 caenorhabdi
6	454	11.0	856	4	Q8TEE1	Q8tee1 homo sapien
7	436.5	10.6	2179	13	Q9DGG9	Q9dgg9 brachydanio
8	436.5	10.6	2179	13	Q9DGG8	Q9dgg8 brachydanio
9	418	10.1	2113	11	Q99MZ6	Q99mz6 mus musculu
10	416	10.1	1204	5	Q8T0A9	Q8t0a9 drosophila
11	416	10.1	2167	5	Q9V3Z6	Q9v3z6 drosophila
12	409.5	9.9	921	11	Q8R0U3	Q8r0u3 mus musculu
13	363.5	8.8	2098	5	P91443	P91443 caenorhabdi
14	344.5	8.3	563	5	Q9VZ45	Q9vz45 drosophila
15	343.5	8.3	548	4	Q8TCK0	Q8tck0 homo sapien
16	304.5	7.4	2424	5	Q9VZ48	Q9vz48 drosophila
17	279.5	6.8	2121	5	Q9NH54	Q9nh54 drosophila
18	278.5	6.7	2129	5	Q9VLZ3	Q9vlz3 drosophila
19	277	6.6	2357	5	Q9U1M8	Q9u1m8 dictyosteli
20	243.5	5.9	578	5	P90536	P90536 dictyosteli
21	213.5	5.2	1260	10	Q9FHN8	Q9fhn8 arabidopsis
22	213.5	5.2	1260	10	Q8VZV4	Q8vzv4 arabidopsis
23	208.5	5.0	1265	10	O24147	O24147 nicotiana t
24	207.5	5.0	1259	10	O22326	O22326 arabidopsis
25	207.5	5.0	1259	10	O23102	O23102 arabidopsis
26	203	4.9	1265	10	Q41460	Q41460 solanum tub
27	199.5	4.8	1261	10	Q39130	Q39130 arabidopsis
28	176.5	4.3	1200	10	Q9FQL7	Q9fql7 zea mays (m
29	167.5	4.1	1094	4	Q9C0H5	Q9c0h5 homo sapien
30	157.5	3.8	796	11	Q8VCE9	Q8vce9 mus musculu
31	153.5	3.7	712	4	Q94893	Q94893 homo sapien
32	152.5	3.7	924	11	Q8R3S0	Q8r3s0 mus musculu
33	152.5	3.7	2062	11	Q9JJY5	Q9jjy5 mus musculu
34	152	3.7	1146	3	Q8X0L3	Q8x0l3 neurospora
35	151.5	3.7	528	4	Q9H7D0	Q9h7d0 homo sapien
36	149	3.6	705	4	Q96B28	Q96b28 homo sapien
37	148.5	3.6	555	12	Q9YMX1	Q9ymx1 lymantria d
38	147.5	3.6	1309	5	Q9VDE9	Q9vde9 drosophila
39	147.5	3.6	3536	5	Q9VZ30	Q9vz30 drosophila
40	147	3.6	2207	5	Q9U0V2	Q9u0v2 leishmania
41	145	3.5	1130	4	Q9C0D6	Q9c0d6 homo sapien
42	145	3.5	1234	4	O75133	O75133 homo sapien
43	144.5	3.5	690	10	Q94D41	Q94d41 oryza sativ
44	142.5	3.5	559	11	Q91W51	Q91w51 mus musculu
45	142.5	3.5	1326	4	Q8TEQ2	Q8teq2 homo sapien

ALIGNMENTS

RESULT 1
Q96JP2
ID Q96JP2 PRELIMINARY; PRT; 1560 AA.
AC Q96JP2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1783 protein (Fragment).
GN KIAA1783.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large Proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058686; BAB47412.1; -.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MyTH4; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00015; IQ; 1
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50002; SH3; 1.
FT NON_TER 1 1
SQ SEQUENCE 1560 AA; 170146 MW; 6A61734C61E4C35A CRC64;

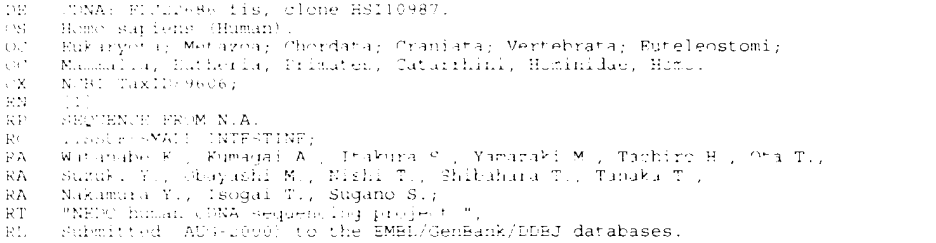
Query Match 53.7%; Score 2217.5; DB 4; Length 1560;
Best Local Similarity 97.5%; Pred. No. 7.6e-156;

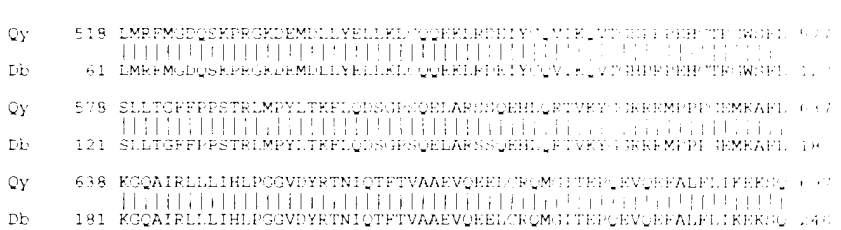
Matches 438; Conservative 0; Mismatches 0; Indels 11; Gaps 1;



RESULT 2
Q9BC14
ID Q9BC14 PRELIMINARY; PRT; 487 AA.
AC Q9BC14;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE cDNA: FLJ20486 fis, clone HSI10987.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Wakamatsu K., Kumagai A., Irakura S., Yamazaki M., Tachiiri H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP001454; BAB15484.1; -.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00784; MyTH4; 1.
DR SMART; SM00295; B41; 1.
DR SMART; SM00139; MyTH4; 1.
DR PROSITE; PS50057; BAND_4_1_3; 1.
DR PROSITE; PS50030; EPM_ENF_1; UNKNOWN_1.
SQ SEQUENCE 487 AA; 55015 MW; 696AAF4001E5BFF9 CRC64;

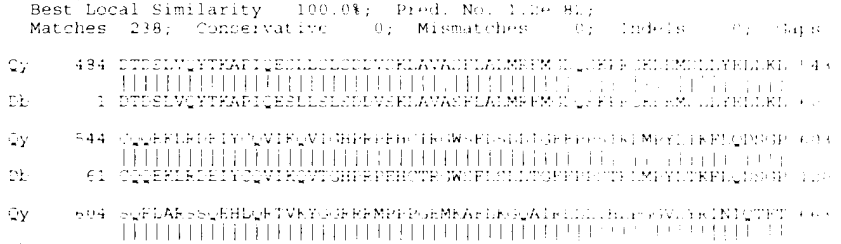
Query Match 32.6%; Score 1346; DB 4; Length 487;
Best Local Similarity 99.6%; Pred. No. 9.2e-92;
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;





RESULT 3
Q8TCJ6
ID Q8TCJ6 PRELIMINARY; PRT; 461 AA.
AC Q8TCJ6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 52.2 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPH NODE;
RA Duesterhoeft A., Lauber J., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL713730; CAD28520.1; -.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 461 AA; 52178 MW; 50EE978DF791F468 CRC64;

Query Match 29.6%; Score 1223; DB 4; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.1e-82;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 4
Q8R1I5
ID Q8R1I5 PRELIMINARY; PRT; 417 AA.
AC Q8R1I5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 46.9 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC024506; AAH24506.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   417 AA;  46896 MW;  0060A1DEE517D985 CRC64;

  Query Match             19.5%;  Score 805.5;  DB 11;  Length 417;
  Best Local Similarity   24.3%;  Pred. No. 1.2e-52;
  Matches  163;  Conservative   19;  Mismatches   22;  Indels    1;  Gaps    1;

Qy     527 KPPGKPPPMFILYELMLQGSEVPEPPIYTQVIKQVTGHPPFFHCTPGWSFLSLLTGFPPP 586
Db       2 KPRGKTPLFLFLYELMLQGSE-LPDPMYCQVIKQVTGHPQPKHCALGWSVLSLFTGFPAP  60

Qy     587 STRLMPYLTKPLLPPSPSPELARSSQEHLQRTVKYGGFPPMPPPGEMKAPLKGQAIPLLL 646
Db      61 STTLMPYVTPPLQPSSPSPELARRSQENLQRTVKYGGFQQLPLPGEMNAPLKGQAVPLLL 120

Qy     647 IHLPGGVDYPTNIQTFTVAAPVQEPLPPQMGITEPQEVQEPALPLIKEKSQLVRPLQPAE 706
Db     121 IHLPGGVDYPTNIQTFTVAGPVIEPLCQQMGITDLPEVQEPALPLIKGEGELVRPLSPHE 180

Qy     707 YLNSVVVDQPVSLHS 721
Db     181 YINPVTTDQPMSLHS 195
```

```
RESULT 5
Q20456
ID   Q20456       PRELIMINARY;      PRT;  2810 AA.
AC   Q20456;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F46C3.3 protein.
GN   F46C3.3
OS   Caenorhabditis elegans
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cottage A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 98068613; PubMed 9851916.
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z66563; CAA91469.1; -.
DR   HSSP; P10587; 1BR2.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 3.
DR   Pfam; PF00784; MyTH4; 2.
DR   ProDom; PD000355; myosin_head; 1
DR   SMART; SM00295; B41; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   PROSITE; PS50096; IQ; 1.
SQ   SEQUENCE   2810 AA;  322526 MW;  27BEE5064B758FB2 CRC64;

  Query Match             13.2%;  Score 545;  DB 5;  Length 2810;
  Best Local Similarity   24.2%;  Pred. No. 4.8e-31;
  Matches  185;  Conservative  127;  Mismatches  294;  Indels  158;  Gaps   24;

Qy      39 NGAHSSPPMLSPSPGKGPPPAVAPRPKAPLQLGPSSSI---------------------- 76
Db    1876 NGNVVPPP---PSGYRAPSPAPTSGDPRGLNRLPQESYTEPAVPGNTKNNSEYLSPHPFNL 1932

Qy      77 ----------KEKQGPLLDLFGQKL--PIAHTPPPPPAPPL--------PLPLDPGTLSA 116
Db    1933 DEQKAVIRHEKETAYNAINMLSDRLRSPVPQPTPPFPPPPVREEWVIPDSVERDPVTQGR 1992

Qy     117 ERRCL-TQPVEDQGVSTQLLAFSGSVCPSYTGTPWKLFLKKEVFTFRENFSHPYYLRLLC 175
Db    1993 IPGSFPFPPLVPQPPPPPVVAEKPAV- KPVKAPWKLTLRKEMFVFHEVLHRLTSQVF 2050

Qy     176 EQILRDTFSESCIRIGQNEPPPKMKDLLGGLEV---DLDSLTTTEDSVKKFIVVAARDNWA 232
Db    2051 AQIVELCKKSYFYRIRQDPRKQVETVLPQYELPPSDLNNLSNTHEPVFTAVIPLAP IWP 2109

Qy     233 NYFSKFFPV----SGESGSDVQLLAVSHFSLPLLFVTSKPSLFLAGLKELSSTEPAFV 288
Db    2110 LYFNQVYEVVEKRPDESVSTI--FAISEHGIPLIVHT- THGLENPLFLAQFFPETIAD 2165

Qy     289 VECKGGSILELSLKSEQ---- - LVLHTAPAPAIEALVELPLNELKEDSGYVIALRSYI 341
Db    2166 VSLEANDILSVHVRHEDEENAYSAVRIKTDQAPQIKETLEKI- - - - SGGWVPKPKFV 2219

Qy     342 TDNCSLLSFHRGDLIKLLPVATLEGWLPGSAGPSGLFPADIVQ-------PAAAPDPS 394
Db    2220 LAPTGR---------ETPFVG----NWLYGKIENPFGFLLAQYVDSTDGDNVPPIRHETS 2266

Qy     395 FSKEQRSGWHKGQLSNGEFGLAPWDPASEPPAHPWSQAHSDDSEATSLSSVAYAFLPDGH 454
Db    2267 EDRDEPVRFFDDEV------------------PFS----------------------SE 2285

Qy     455 SYTMQEFAPPYPPPSQALLGQTDGGAAGKDTDSLVQYTKAPIQESLLS-LSDDVGKLAVA 513
Db    2286 RYTMIDFATKYPPKPKDKKKQETW--AWEDISQIVRFSEKPISQSLLADLGNEECPYAVE 2343

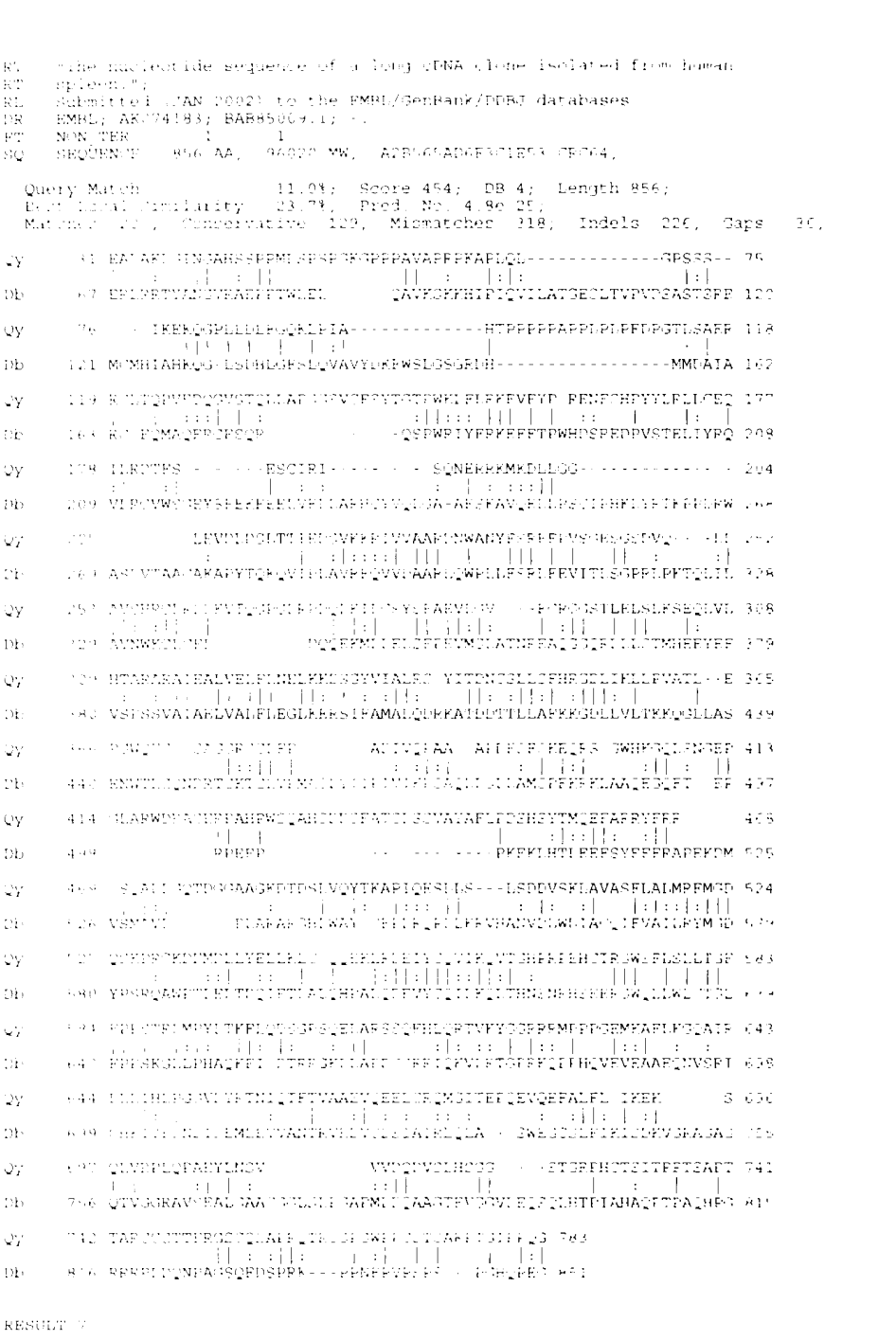
Qy     514 SFLALMRFMGDQSKPRGKDEMDLLYELLKLC QGFKLPDFIYCQVIKQVTGH--PRPEHC 570
Db    2344 TFHAIMKFMGDEPLKKSESMTDVVFKVLLICHRQPTLRLEVIQQLIRQTTLNLLPPPNSA 2403

Qy     571 TRGWSFLSLLDGFFPPSTPLMPYLTKPLQDSG-----PSQELARSSQEHLQRTVKYGGRR 625
Db    2404 LRAWRLLTIITAYFPSSLTLKPYVLQYLGDNADEWQKPFHGIAPIQVINMLQTFFTVAAE 2463

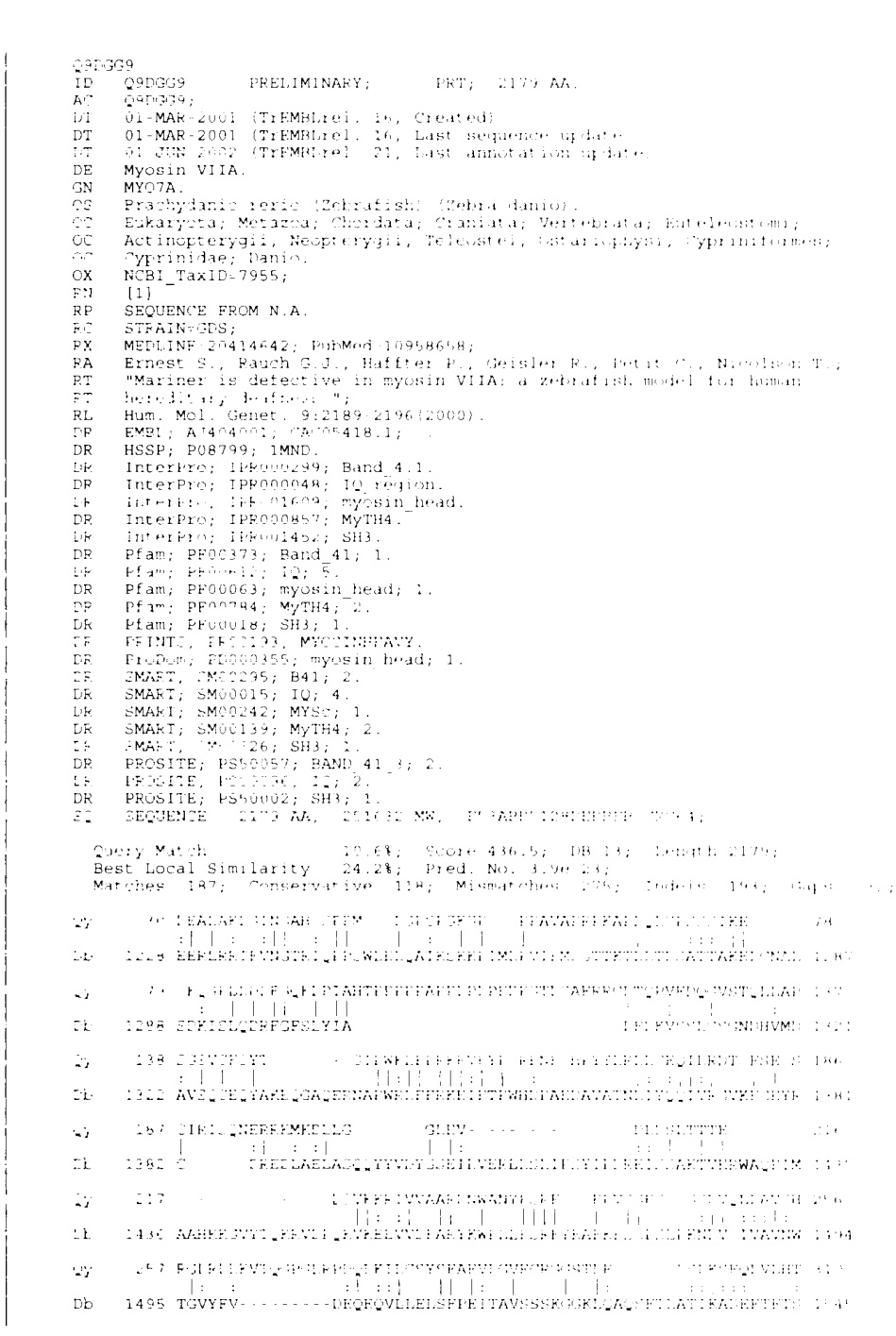
Qy     626 RMPPPGEMKAPLKGQAIRLLLIHLPGGVDYR----------TNIQTFTVAAEVQEEL 672
Db    2464 VLLNALEVQQITVGSIVTLYPGESTDGQQLRRQAFYISKDENVGQTLFFITYKFFMIQEL 2523

Qy     673 CRQMGITEPQEVQEPALFLIKEKSQLVRPLQPAEYLNSVVVDQD 716
Db    2524 CNLLNVRSLHEQQEFSLGYTVGKDPHLNYCFNDNTLMDIITESE 2567
```

```
RESULT 6
Q8TEE1
ID   Q8TEE1       PRELIMINARY;      PRT;   856 AA.
AC   Q8TEE1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FLJ00256 protein (Fragment).
GN   FLJ00256.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPLEEN;
RA   Jikuya H., Takano J., Nomura N., Fikuoka F., Nagase T., Ohara O.;
```



RESULT 7



```
Qy      311 ARARAIEALVELFLNELKKDSGYVIALR---SYITDNCSLLSFHRGDLIKL---LPVATL 364
            |  |  ||  ||  |:| | :|:||:   |   |: : ||| :|||| |         :
Db     1546 NNAEDIRDLVVTFLEGLPKPSKFVVALQDNPSPAADDSTFLSFLKGDLIVLDQDTGEQVM 1605

Qy      365 EPGWQFGS--AGGRSGLFPAD------IVQPA -AAPDFSFSKEQPSGWHKGQLSNGEP 413
            || |:     : | ||||       .|.|        : . :||
Db     1606 TSGWAHGTNDRTKQPGDFPADCVYVLPTVVRPPHDVVAFVTMTPDQR------------ 1652

Qy      414 GLARWDRASERPAHPWSQAHSDDSEATSLSSVAYAFLPDSHSYTMQEFARRYFR------ 467
                : | | :|              .|.      |   ||:-||:  |||
Db     1653 ------QESLRISH-----------VPALTETEERVKP----YTLEEFSYDYFRPPPKHT 1691

Qy      468 RSQALLGQTDGGAAGKDTDSLVQYTKAPIQESLLSL---SDDVSKLAVASFLALMRFMGD 524
            |. :: . |          | |   |: ||:|:||       :::|. |  :|.|.|::|||
Db     1692 LSRVMITKNRG-- - -KDKLWCCTPFPIYQALLKVCGGHEFLSQEACMAFIAVMKYMGD 1745

Qy      525 QSKPRGKDEMDLLYELLK-LCQQEKLRDEIYCQVIKQVTGHPPPEHCTPGWSFLSLLTGF 583
            |  :  :|  : :|          | ||:          :|  | | |
Db     1746 YPSKRTRSVNELTDQIFEGALFAAFLPDEIYCQILKQLTENHIKYSEEKGWELIWLCVGL 1805

Qy      584 PPPSTRLMPYLTKFLQDSGPSQELAKSSQEHLQRTVKYGGFPMPPPGEMKAFLKGQAIR 643
            |||| |:|:: :||| |      ||    . ||. :: | |:  |   |:::|
Db     1806 FPPSNVLLPHVQRFLQ-SKKHHPLALDCMQFLQKALPNGSPKYPPHLVEVEAIQHKTTQI 1864

Qy      644 LLLIHLPGGVDYRTNIQTFTVAAEVQEELCRQMGITEPQEVQEFALFL-IKEK 695
            :: |  |     ::: | |  :          :: :  |   : |:||: | :|
Db     1865 FHKVYFPDDTDEAFEVESSTKAFDFQLNISSRLLLKTP--EGFSLFVKISDK 1914
```

RESULT 8
Q9DGG8
ID   Q9DGG8   PRELIMINARY;   PRT; 2173 AA.
AC   Q9DGG8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Myosin VIIA.
GN   MYO7A.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TUEBINGEN;
RX   MEDLINE=20414642; PubMed=10958658;
RA   Ernest S., Rauch G.J., Haffter P., Geisler R., Petit C., Nicolson T.;
RT   "Mariner is defective in myosin VIIA: a zebrafish model for human
RT   hereditary deafness.";
RL   Hum. Mol. Genet. 9:2199-2206(2000).
DR   EMBL; AJ404002; CAC05419.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50002; SH3; 1.

```
SQ   SEQUENCE   2179 AA;  251655 MW;  FF440D1EFD588F76 CRC64;

  Query Match           10.6%;  Score 436.5;  DB 13;  Length 2179;
  Best Local Similarity   24.2%;  Pred. No. 3.9e-23;
  Matches  187;  Conservative  118;  Mismatches  275;  Indels  193;  Gaps   32;

Qy       30 DEALAKLGINGAHSSPPM---LSPSPGKGI-- FPAVAFPPKAPQQLGPSSSIKE----- 78
            :| | :  :||  : ||    | :  | |   |                 |  ::: ||
Db     1228 EEPLPPTPVNGTRTQPPSWLELQATKSKEPIMLPVTFMDGTTKTLLTPSATTAKELCNAL 1287

Qy       79  KQGPLLDLPGQKLPIAHTPPFPAPFLPLPEDPGTLSAPPRCLTQPVEDQSVSTQDLAP 137
             :  | | ||  | ||                                 |   |   :
Db     1288 SDKISLQDRFGPSLYIA---------------------- LPDKVSSLGSGNDHVMD 1321

Qy      138 SGSVCFSYT------- GTLWKLPLPFEVEYP PENFSHFYYLPLLCPQILRDT-PSE-S 186
            . | |  |                ||:|| |||:| |  :                |: :|||:|  | |
Db     1322 AVSQCEQYAKEQGAQEPNAPWPLPFPKEIFIPWHDPAELAVATNLIYCGTVFGVFPGPYP 1381

Qy      187 GTPISQNFPPKMKDLLG-------GLEV  - - - - - - - - - - -DLDSLTTTE---  216
            |      :|  : :|           | ||:                    :: |  | |
Db     1382 C------DREDLAELASQQYYVDYGSEILVEKLLSLIFSYIIPFEISCAFTVFPWAQFIM 1435

Qy      217 ---------------DSVKKRIVVAARDIWANTFSFF- -FFTSLSL  TSTVQILAVSH 256
                          ||: :|  ||  |  ||||   |  || |  .|| ::|||:
Db     1436 AAHKKGVYTQKRVDPQKVKEEVVDFARYKWPLLFSEFYEAFFPSSLGLFFNDV IVAVNW 1494

Qy      257 RGLRLLKVTQGPGLRFDQLKILCSYSPAEVLGVECFGSILE-      LSLKSPQLVLHT 310
            |  |  |  |  |  :| ::|  || |: |  |  |  |:          ::|:::  :
Db     1495 TGVYFV---------DEQEQVLLELSFLEITAVSSSFSGFFLRALGFTLATIFADFIFTFT 1545

Qy      311 ARARAIEALVELFLNELKKDSGYVIALR---SYITDNCSLLSFHRGDLIKL---LPVATL 364
            |  |  ||  ||  |:| | :|:||:   |   |: : ||| :|||| |         :
Db     1546 NNAEDIRDLVVTFLEGLPKPSKFVVALQDNPSPAADDSTFLSFLKGDLIVLDQDTGEQVM 1605

Qy      365 EPGWQFGS--AGGRSGLFPAD-------IVQPA -AAPDFSFSKEQPSGWHKGQLSNGEP 413
            ||  |:     : | ||||         :|:|         : : :||
Db     1606 TSGWAHGTNDRTKQPGDFPADCVYVLPTVVRPPHDVVAFVTMTPDQR---------- - 1652

Qy      414 GLARWDRASERPAHPWSQAHSDDSEATSLSSVAYAFLPDSHSYTMQEFARRYFR-- -   467
                : | | :|                :|:       |    ||:::||:  |||
Db     1653 ------QESLRISH------- ---VPALTETEERVKP - -YTLEEFSYDYFRPPPKHT 1691

Qy      468 RSQALLGQTDGGAAGKDTDSLVQYTKAPIQESLLSL  --SDDVSKLAVASFLALMRFMGD 524
            |. :: . |          | |   |:  |. . |:|:|||:--||:| :          :||  | | |
Db     1692 LSRVMITKNRG-- - -KDKLWCCTPFPIYQALLKVCGGHEFLSQEACMAFIAVMKYMGD 1745

Qy      525 QSKPRGKDEMDLLYELLK-LCQQEKLRDEIYCQVIKQVTGHPPPEHCTPGWSFLSLLTGF 583
            |  :  :|  : :|          | ||:          :|  | | |
Db     1746 YPSKRTRSVNELTDQIFEGALFAAFLPDEIYCQILKQLTENHIKYSEEKGWELIWLCVGL 1805

Qy      584 PPPSTRLMPYLTKFLQDSGPSQELAKSSQEHLQRTVKYGGFPMPPPGEMKAFLKGQAIR 643
            |||| |:|:: :||| |      ||    . ||. :: | |:  |   |:::|
Db     1806 FPPSNVLLPHVQRFLQ-SKKHHPLALDCMQFLQKALPNGSPKYPPHLVEVEAIQHKTTQI 1864

Qy      644 LLLIHLPGGVDYRTNIQTFTVAAEVQEELCRQMGITEPQEVQEFALFL-IKEK 695
            :: |  |     ::: | |  :          :: :  |   : |:||: | :|
Db     1865 FHKVYFPDDTDEAFEVESSTKAFDFQLNISSRLLLKTP--EGFSLFVKISDK 1914
```

RESULT 9
Q99MZ6
ID   Q99MZ6   PRELIMINARY;   PRT; 2113 AA.
AC   Q99MZ6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Myosin-VIIb.
GN   MYO7B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21295050; PubMed=11401444;
RA   Chen Z.-Y., Hasson T., Zhang D.-S., Schwender B.J., Derfler B.H.,
RA   Mooseker M.S., Corey D.P.;
RT   "Myosin-VIIb, a novel unconventional myosin, is a constituent of
RT   microvilli in transporting epithelia.";
RL   Genomics 72:285-296(2001).
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AF242411; AAK28339.1; -.
DR   HSSP; P19799; 1MND.
DR   MGD; MGI:107709; Myo7b.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000871; Beta_lactamase_A.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR000508; SigPTase.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00660; BAND_41_3; 2.
DR   PROSITE; PS00146; BETA_LACTAMASE_A; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS00761; SPASE_I_3; UNKNOWN_1.
KW   SH3 domain.
SQ   SEQUENCE   2113 AA;  240795 MW;  EC8B0916 3D0DCA69 CRC64;

  Query Match             10.1%;  Score 418;  DB 11;  Length 2113;
  Best Local Similarity   24.5%;  Pred. No. 9.9e-22;
  Matches  179;  Conservative  101;  Mismatches  248;  Indels  202;  Gaps   30;

Qy         PKPEALAKLGINNAHSPPPMLPP              SPGKSPPP       AVAPPPKAPLQI 70
Db    1120 PPSPPPMPEIIDELTLQ...ETERTVEAFFETFAHTVEAFFPTWLEIQAVKEPPHIFIQV 1138

Qy      71 ...............................EEQGPLIHPGQFLPIA-------------HTPP 98
Db    1199 ILATGRSLTISVDSASTSREICQHVAQKQGLRDNLGFSLQVAVYDKFWSLGSGLDH   1254

Qy      99 PPPAPPLPLPPPPGTPCAPPPGITQPVEPQGVSTQLLAPSGSVCPSYTGTPWLFLPKPEV 158
Db    1255 ------------LMDAVAQC                FQLAPEPGE    SQRLAPWRIYFKPEF 1286

Qy     159 ...........................................................  184
Db    1287 ...........................................................  1346

Qy     185      ..........................FVTLPLTTTTEPSVEEEIVVAAPPNWANYF 235
Db    1347 ...........................LLAASAPALYTTSPATPLAVPPQTVPAAPLLWPLLP 1406

Qy     236 ...........................LLAVSHPGLPLLFVTQGPGLPPGQ.FIIPGYSPAEVLGV 289
Db    1406 SRLPPPT...................VLAINWPPMYEP          PQPEPTLLGLSPAEVMGLV 1455

Qy     290 ...........................................................LITLPLTGLPPPQLVLHTARAPAIPALVEIPINLPPPSSYVIALPS YITPNT 345
Db    1456 ANRQARGGRLLLALQLEEYEPVSPLSVAIAEMVALPQQGLPEPSVFAMALQPPRAIEDI 1515

Qy     346 SLISPHPSPLIKLLPVATL  EPGWQTD  IASSPSLPLA ......LLVLPAA- ALD 392
Db    1516 TLLPPPPSPLLLITPKQSLLASENWALSQMRPGRIGLVPTALLYIIPSVTPPSTQLLSL 1575

Qy     393 FSPSKEQFSGWHEQQLSNGEPSLARWPPASERPAHPWSLAH......FAT...SVAYAPLL 452
Db    1576 LAMSPEKRK-LAAQEVFALPPPLP  PQLTHSP                              16..

Qy     453 SHSYTMQEFARRYFRRSQALLGQTHQGAA   ..............................VSF 509
Db    1636 ..........LLPPSSLPPPAFLP   .....AAM.MAP.......WAY............VHL KAE 16..

Qy     510 LAVAS    FLALMRPMGDQSKPPGKDPMPLLYELLE..........PLPPPLY....VIPLVT HH 5..
Db    1660 LRDAACQIFLAILKYTGDYPSRQSWHSLELTDQMPSLALQLIALQLELYCQILKQLTHNS 1719

Qy     566 KPPHQIPGWSPLSLLTGPPPPSIPIMPPLLKPP..............................RTVEY..HH 6..
Db    1720 IRPSEEPAWQLIWLCTGLPPPGKILLPHAQPPI.........PPLAL..............HRVLPV.PPP 1778

Qy     626 EMPPPGEMKA 635
Db    1779 QPPHDVEVKA 1788
```

```
RESULT 10
Q8T0A9
ID   Q8T0A9      PRELIMINARY;      PRT;  1254 AA.
AC   Q8T0A9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   LD14917p.
GN   CK.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY069438; AAL39583.1; -.
SQ   SEQUENCE   1254 AA;  144367 MW;  806DEC6DB7ADL3F4 CRC64;

  Query Match             10.1%;  Score 416;  DB 4;  Length 1254;
  Best Local Similarity   23.3%;  Pred. No. 5.7e-22;
  Matches  183;  Conservative  115;  Mismatches  283;  Indels  206;  Gaps   29;

Qy      20 KAPLKKIDEE ...DEALAKLGINNAHSPPPMLPP...........AVAP       PPKAP ..
Db     292 KAPIKK...YAPY...............IENN....PPSWLEIQAIKEPPIMIPITPMPYNTKTL 3..

Qy      ..........................................................................  11.
Db     352 LAPSATTAPELTNILSPKIILK-      IQPGPSLYIA                        3..

Qy     116 AEEEL..........................................................YPPRNPTH 1..
Db     384       .............................................................WHPPTH 4..

Qy     168 PYY ...................................NPPPKMPDL-                   ..1
Db     433 ...VA........YLLVVPS       VHLSLPPLDPPELAMIAALLYP............PTTI 4..

Qy     202 ...... LGGLEVDLPSLTT             TEP      SVEPPIVVAAPLNWANYPPR 2..
Db     492 INPIELDGLLGVLPAIEEWAALVLLAYPP..........IAPLKI............YAPYKWPLLPP 5..

Qy     238 PPPVQSE...LVL    LLAV...........................................VPKRG 2..
```

```
Db    553 FYEAYRNSGPNLPFNDVIIAVNWTGVYVV---------DDQFQVLLELSFPEITAVSSQK 603

Qy    294 GS-----TLELS-LKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALPSY--ITDNC 345

Db    604 TNKVFTQTFSLSTVRGEEFTFQSFNAEDIPDLVVYFLDGLFKPSKYYVIALQDYPAPSDGT 663

Qy    346 SLLSFHRGDLIKL---LPVATLEPGWQFG--SAGGPSGLFPADIVQPAAAPDFS----- 394

Db    664 SFLSFFKGDLIILEDESCGESVLNNGWCIGPCDPSQEGDFPAETVY--VLPTLSKPPQD 721

Qy    395   PSKPCPSGWHKGLLSNGFFRLAPWDPASEPPAHPWSQAHSDDSEATSLSSVAYAFL 450

Db    722 ILALPNIEPA---HHGR---------- . ---------RLSMASNGGAVE 747

Qy    451 PDGHSYTMQEFARPYFRRSQALLGQTDGGAAGKDTDSLVQYTKAPIQESL---LSLSDDV 507

Db    748 PRDRPHTLMEYALDHFRLPFEFTMSPTLTLSSPRSEELWRYSFDPIVAELLRKLQSKEEF 807

Qy    508 SKLAVASFLALMSFMGD- QSKPP-GKDEMDLLYELLKLCQQEKLRDEIYCQVIKQVTGH 564

Db    808 AEEACFAFAAILKYMGLDSPSFPPFMGNEITDHIFD--GPLKHEILRDEIYCQLMKQLTDN 865

Qy    565 PPPEHCTPGWSFLSLLTGFFPPSTRLMPYLTKFLQDSGPSQELARSSQEHLQRTVKYGGR 624

Db    866 RNRMSEERGWFLMWLATGLFACSQGLLKELLNFLRTR- PHPISQDSMHPLQKTIPHGQR 923

Qy    625 PPMPPPGPMKAFLKGQAIPLLIHLPGGVDYRTNIQTFTVAAEVQEELCRQMGITEPQEV 684

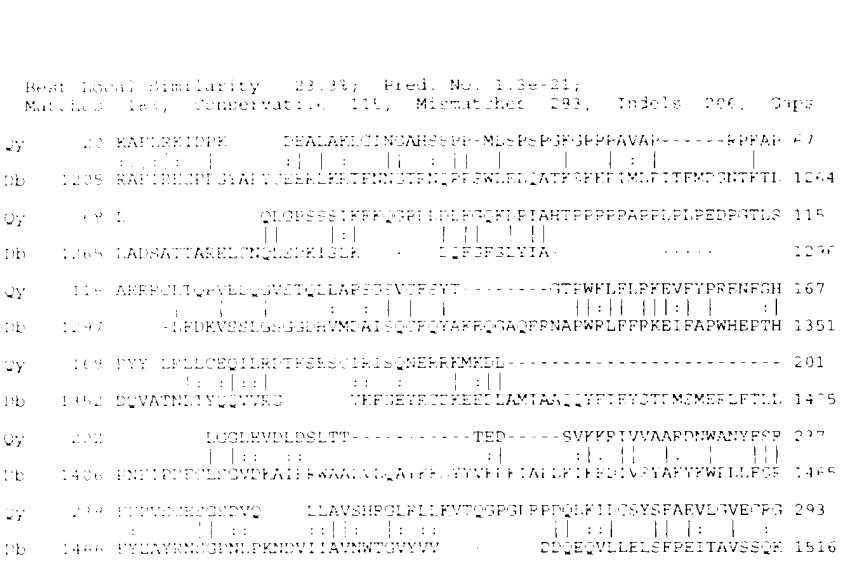
Db    924 KYPPHQVEVEAIQHKTTQIPHFVYFPDDTDEAFEVDSSTRAKDFCNNISQPLSI---PTS 980

Qy    685 QEFALFL 691

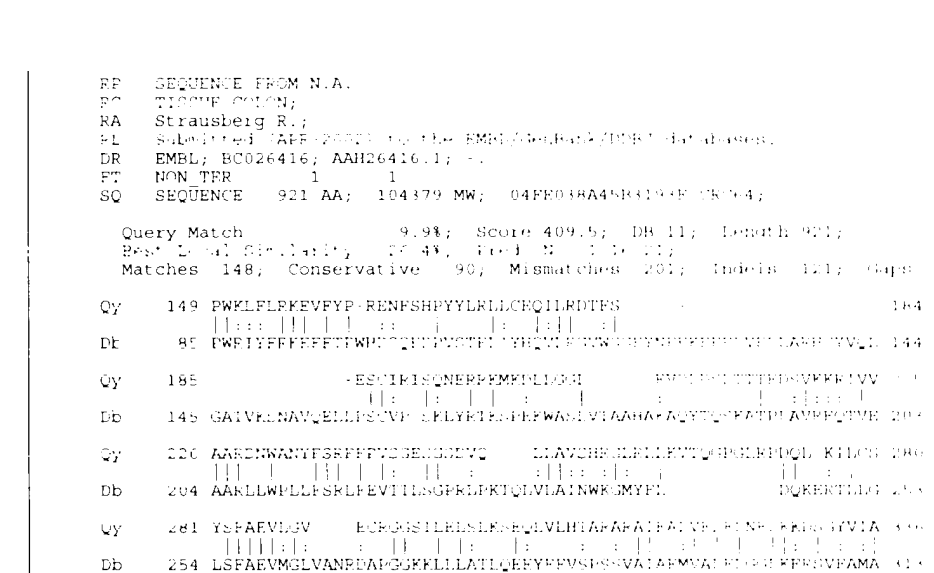
Db    981 EGFSLFV 987


RESULT 11
Q9V3Z6
ID   Q9V3Z6      PRELIMINARY;      PRT;  2167 AA.
AC   Q9V3Z6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CK gene product.
GN   CK OR CG7595.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera, Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle P.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.,
RT   "An exploration of the sequence of a 2.9 Mb region of the genome of
RT   Drosophila melanogaster: the Adh region";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang P.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AE003646; AAF53435.1; -.
DR   EMBL; AE003412; AAF44915.1; -.
DR   HSSP; P08733; 1MND.
DR   FlyBase; FBgn0000317; ck.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR000567; SBP_bac_1.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00612; IQ; 4.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS50096; IQ; 1.
DR   PROSITE; PS01037; SBP_BACTERIAL_1; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain.
SQ   SEQUENCE   2167 AA;  250307 MW;  3C57B34ADFD88A42 CRC64;

  Query Match             10.1%;  Score 416,  DB 5;  Length 2167;
```



RESULT 12
Q8R0U3
ID Q8R0U3 PRELIMINARY; PRT; 921 AA.
AC Q8R0U3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to myosin VIIb (Fragment).
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxID=10090;
RN [1]



RESULT 13
P91443
ID P91443 PRELIMINARY; PRT; 2098 AA.
AC P91443;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE T10H10.1 protein.
GN T10H10.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150218; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

```
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston P.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 369:32-38(1994)
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Nelson J., Langston Y.;
RT   "The sequence of C. elegans cosmid T10H10.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80848; AAB37988.1; -.
DR   HSSP; P09493; 1MND.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF00612; IQ; 4.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00314; RA; 1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS50096; IQ; 1.
SQ   SEQUENCE   2098 AA;  239783 MW;  A972E51FC54B73BB CRC64;

  Query Match             8.8%;  Score 363.5;  DB 5,  Length 2098;
  Best Local Similarity   21.7%;  Pred. No. 9.9e-18;
  Matches  176;  Conservative  125;  Mismatches  320;  Indels  191;  Gaps   30;

Qy      22 FLRKIDPF-  --  DPALAKLGINGA-HSSPPMLSPSPGKGPPPAV------------ 60
           |:|:  |       ..| . .:||  |  |  :     |   | |
Db    1187 FIRERSFASTDYRPYYIPIFLRFTWNSTPHQPFSYVELQANKSQFPVVLAVTFMDGSVKT 1186

Qy      61 --AFSPPAPLQLGPSGSIKFKASLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAEP 118
             |      :|     : ||| | | : ||  | ||
Db    1187 LCADSATTAAEL--CKQLAEKVG-LTNSPGFSLYIA  ---------------------- 1219

Qy     119 RCLTQPVEDAGVSTQLLAPSGSVCFSYT-  --  --GTPWKLFLPKEVFYP-PENFSHPY 169
           |   |  |  |  :  | | |  |  |  |  |  ||.|| |||-| | :   :
Db    1220 - LFLKYGSLSSSTPHYMDAISQLPQYAFEQTRQERNAFWFLFFPKEIFSPWHDPPDPV 1277

Qy     170 YLRLLCEQILR-----  --  --  ------ ----DTFSESCIPISQNERRKMKDL-LG 203
             |:  |..|                        |  .   ::  |    : | :
Db    1278 STNLIYQQVIFSIEYSEYPDKDEELAAICALVYIDESTMDVNKLFNRLFSYLPDFEMS 1337

Qy     204 GLEVDLDSLTTT--  --  --  ----EDSVKKRIVVAARDNWANYFSRFFPVSGESGSD 248
           | |: |:   | |                :  ||: :|  |:  |   ||||:     :|
Db    1338 GKEMALEKWTQTIMHLVPFFFTGRLPSQIEVFENVVSVAKTKWPLLFSPFYFALKFAGPP 1397

Qy     249 VQ--  LIAVSHPTLPLLFVTQN SLPPDQLFILCSYSFAEVLGVECRQG -STLEL -- 299
           :     :.|:  |..  |                |:  :.  :|: | :       |  ||
Db    1398 LPKNEVIIAVNWTGVYVV---------DDREHVMLEFSFPEISTAYYGKGKRSTTDTCTV 1448

Qy     300  SLFSEQLVIHTAPAPAIEALVELFLNELKKDSGYVIALPSYITD-NCSLLSFHRGDLIK 357
            ::  ::    :  |  |  |. .||  |||  | |::|:|  |   :  | | :|||:
Db    1449 RTVVGDEYTFQSPNADDITNLIVMPLEGLKFFSPYLVAIFSQKGDEKINFLEFFFGDLLI 1508

Qy     358 LLPVATLEPGWQFGSAGGRGGLFPADIVQFAAAPDFSFSKEQPSGWHKGQLSNGFPGLAR 417
           |:   |              : |   :|                          ||:  |   |
Db    1509 LVNEFT------------GNTLLTESVV-------------------KGE--NSRTCLFG 1545

Qy     418 WDRASE--  ----PFAHPWSQAHSDDSEAT--  --  --SLSSVAYAFLPDSHSYTMQEF 461
             ||           :|:   |    | : :          .. | |    ::  ||:: |
Db    1546 LIRAENVYVLFTLVKFSKNTLQIFPFDMDLSLDLFNNNFQVTVVPY----NAFPYTLENF 1601

Qy     462 ARRYFRRSQALLGQTDGGAAGKDTDSLVQ---YTKAFIQESLLSL---SDDVSKLAVASF 515
           |  |      :|          :  :| ::   ...  | : ||       :|  : |:  |
Db    1602 AEDNFNSQVKPVGSQISLMTLPKFESQIECWRFSREHIDQPLLKKLNGFEDACRGAIEIF 1661

Qy     516 LALMRFMGDQSKFRGKDEMDLLYELLKL CQQEKLPDFIYCQVIKQVTGHPRPEHCTFGW 574
           |:|::|||.   | :   |  : ||    | ||||:|||::||:| :|      ||:
Db    1662 AAIMKYMGDFPSKFSRLGTHLTDHIFKLPISMEALRDELYTQLVEQLTLNPSIMSEEPGW 1721

Qy     575 SFLSLLTGFPPPSTRLMFYLTKFLQDSGPQCELARSSQEHLQETVEYSGPPPPMPPPGFMK 634
           | : |: | ||  |   .. ||: | |  .:|  |  :|  | | | |:   |::
Db    1722 ELLWMATGLPAPSAALAKEISHFLK SRP HPIALDCQNPMQKLAVGGSPKYPPHLVEVE 1781

Qy     635 AFLKGQAIELLLIHEHGGVDYETNIQQFTVAAEVQEELGFQMKITSFQEVGPEFALFL IK 693
           |          :  |  :  |  :  | | |  .  .. ..|.  :      :|:|: ||
Db    1782 AIQHKITQIFHKVFFFDNTDEAIEVDSATFARDFCHKIGYPLGL-- KSSDGFSIFVKIK 1836

Qy     694 EKSQLVEPLQPAEYLNSVVVDQDVSLHSGGST 725
           :|  |  :  :|:    |       :|  :|
Db    1837 DK- -VLAVPESEFFFDYVPSLSDWVHTNHAT 1855


RESULT 14
Q9VZ45
ID   Q9VZ45      PRELIMINARY;      PRT;   563 AA.
AC   Q9VZ45;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG2136 protein.
GN   CG2136.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera, Endopterygota, Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

```
RA   [illegible] D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacek J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003484; AAF47983.1; -.
DR   FlyBase; FBgn0030255; CG2136.
DR   InterPro; IPR000857; MyTH4
DR   Pfam; PF00784; MyTH4; 1.
DR   SMART; SM00139; MyTH4; 1.
SQ   SEQUENCE   563 AA;  63811 MW;  353FBC7D09E0D9A0 CRC64;

  Query Match             9.3%;  Score 344.5;  DB 5,  Length 563;
  Best Local Similarity  31.0%;  Pred. No. 3.6e-17,
  Matches   92;  Conservative   51;  Mismatches  135;  Indels   19;  Gaps    6;

Qy     491 [illegible] 504
           :: :|: ::|| : | | |  :|| :   |:::  |: || |
Db      62 PSVIRYSLLQFAMQHFKNDQ LRIADFHHERHQSAANPSYAFLVKWQGHAIRLPLDRLP 89

Qy     505 [illegible] 563
           : :: :: |  ::|: ||       |:  :| :|  |  :   ||||:|||::|| |
Db      90 NCLAPLALECFDCILRYCGEIPLDFDLTEVRCVTTVLMHCHKYLALRDEVTQQLMKQTTA 149

Qy     564 H--PPPEHCTPGWSFLSLLIGPPPPSIRLMPYLTKFLQDSGPSQE-----LAKSSQEHLQ 616
           :  | ::   | |  ||:|  :|  |  | ||| : |  :   :        |     :|:
Db     150 NRSPVPDSSQRAWRLLSILAAYFGCSDALRPYLMEHLTSAASDFRFSCHGTAAVCLTNLP 209

Qy     617 RTVKYQSPREMIIEIEMKAELRIQATRILLIMEFGDVDYFTNIQIFTTVAAFVLFLPLM 676
           ::  : |||: :|  |:  |  |:: |  :  ||||  :  | :  || |:|  |||  :
Db     210 RTAPVQDPEVTEVFETTAVLAGLAPPLIRIISAPPVVDPRNTVVALVIAELVALLPAL 269

Qy     677 [illegible] 733
           ::   | |||:|: | :      ||   ||:  |  :   | ||       | |:
Db     270 GVPSPAPQQEPSLYCIVQGIAFIMPLAADEYILDVTTE---LLKSGQPFYLIPCPSV 323


RESULT 15
Q8TCK0
ID   Q8TCK0      PRELIMINARY;      PRT;   548 AA.
AC   Q8TCK0;
DT   01 JUN 2002 (TrEMBLrel. 21, Created)
DT   01 JUN 2002 (TrEMBLrel. 21, Last sequence update)
DT   01 JUN 2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 60.5 kDa protein.
GN   DKFZp434O1119.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL713794; CAD28548.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   548 AA;  60454 MW;  274F57AA7B86E804 CRC64;

  Query Match             9.3%;  Score 343.5;  DB 4;  Length 548;
  Best Local Similarity  36.6%;  Pred. No. 4.2e-17;
  Matches   71;  Conservative   46;  Mismatches   67;  Indels   11;  Gaps    4;

Qy     534 MILLYEL--LRLC QQEKLPDEIYCQVIKQVTGH--PEPEHCTPGWSFLSLLTGFFPPS 587
           : ::  |   |:||   | :||| ||||:|||:: :   : : | |||  | ::| :   |
Db      37 MSLVCNLPAPLQLCSDHEYMRDEVYCQVVELITNNSVPLPG-GPSGLLTYTAYHQGS 150

Qy     588 TPLMPYLTKFLQDSG--  --PSQELAKSSQEHLQRTVKYQSPREMIIEIEMKAELRIQA 645
           |  |:|||:|||       | | :|:: ::||:|:::| :  :;   |::: | :::
Db     151 EVLRPHLTEFLQVSRTPGIGPPQGILAPACPQNQQRTLRPQGRILELPSSIELRAMLAGRG 210

Qy     646 PLILIHLPGGVDYFTNIQTFTVAAFVQFFLPLMHITPPQPVQPPA-PIIPEPSQLVPPL 704
           :  | |||::    :| ||| :| ||:| :| :|  :  :: :::: : : | :
Db     211 PFQPLPLPGGLPPHPPRTTTVALQVVPFLVALMALTPPASPLPSTPTPNPVQPVPPL 270

Qy     705 QPAFYINSVVVDQD 716
           |:  |  : :
Db     271 SRPAYIIDVASPMP 284


Search completed: January 10, 2003, 02:22:59
Job time : 64.9672 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein protein search, using sw model

Run on: January 10, 2003, 02:19:59 , Search time 25.7906 Seconds
(without alignments)
896.699 Million cell updates/sec

Title: US-09-803-126-6
Perfect score: 4129
Sequence: 1 MYQSRPGPVFVPVQESEPP ... .WPPQSTSARPTGIPPQGFTC 786

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

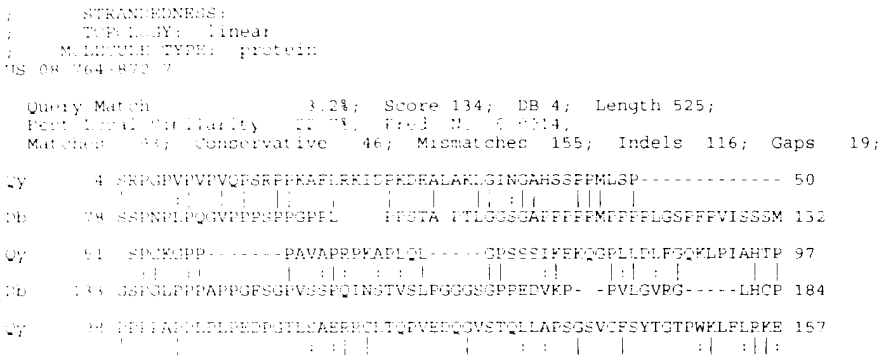
Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	3.2	525	4	US-08-764-870-7	Sequence 7, Appli
2	134	3.2	525	4	US-08-980-115-7	Sequence 7, Appli
3	133	3.2	533	1	US-07-952-800-2	Sequence 2, Appli
4	130.5	3.2	983	4	US-09-311-236-2	Sequence 2, Appli
5	124	3.0	1618	1	US-07-853-913-4	Sequence 4, Appli
6	119.5	2.9	788	4	US-08-630-915A-30	Sequence 30, Appl
7	119	2.9	538	4	US-09-309-572-12	Sequence 12, Appl
8	119	2.9	1737	4	US-09-309-572-13	Sequence 13, Appl
9	118.5	2.9	174	4	US-08-818-112-143	Sequence 143, App
10	118.5	2.9	174	4	US-08-818-111-138	Sequence 138, App
11	118.5	2.9	174	4	US-09-056-556-143	Sequence 143, App
12	118.5	2.9	174	4	US-09-072-596-138	Sequence 138, App
13	118	2.9	214	1	US-08-217-327-4	Sequence 4, Appli
14	118	2.9	1248	2	US-09-080-897-2	Sequence 2, Appli
15	118	2.9	1248	4	US-09-323-735-2	Sequence 2, Appli
16	117.5	2.8	1162	2	US-08-728-323A-2	Sequence 2, Appli
17	117.5	2.8	1162	4	US-09-298-568-2	Sequence 2, Appli
18	117	2.8	538	4	US-09-370-368-9	Sequence 9, Appli
19	117	2.8	1315	3	US-08-899-595-3	Sequence 3, Appli
20	116	2.8	446	1	US-07-952-800-4	Sequence 4, Appli
21	116	2.8	659	4	US-08-671-354-2	Sequence 2, Appli
22	116	2.8	1017	4	US-09-600-776-6	Sequence 6, Appli
23	116	2.8	1805	1	US-07-853-913-2	Sequence 2, Appli
24	115.5	2.8	1255	2	US-09-080-897-4	Sequence 4, Appli
25	115.5	2.8	1255	3	US-08-899-595-1	Sequence 1, Appli
26	115.5	2.8	1255	4	US-09-323-735-4	Sequence 4, Appli
27	115.5	2.8	1871	2	US-08-694-869-1	Sequence 1, Appli
28	115.5	2.8	1871	3	US-09-349-546-1	Sequence 1, Appli
29	114.5	2.8	659	4	US-09-562-737-16	Sequence 16, Appl
30	114.5	2.8	2972	4	US-09-579-181-2	Sequence 2, Appli
31	114.5	2.8	3118	4	US-09-579-181-1	Sequence 1, Appli
32	114	2.8	1274	4	US-09-095-443-2	Sequence 2, Appli
33	113.5	2.7	267	2	US-08-301-162-16	Sequence 16, Appl
34	113.5	2.7	267	4	US-09-461-240-16	Sequence 16, Appl
35	113.5	2.7	267	4	US-09-968-927-16	Sequence 16, Appl
36	113.5	2.7	380	2	US-09-026-587-4	Sequence 4, Appli
37	113.5	2.7	380	2	US-09-227-420-4	Sequence 4, Appli
38	113	2.7	902	1	US-08-396-479B-6	Sequence 6, Appli
39	113	2.7	902	1	US-08-818-823-6	Sequence 6, Appli
40	112.5	2.7	2414	1	US-08-227-536-2	Sequence 2, Appli
41	112.5	2.7	2414	5	PCT-US95-04682-2	Sequence 2, Appli
42	111.5	2.7	398	4	US-09-303-064-54	Sequence 54, Appl
43	111.5	2.7	398	4	US-09-086-503-54	Sequence 54, Appl
44	111	2.7	570	2	US-08-967-364-1	Sequence 1, Appli
45	111	2.7	570	3	US-09-368-408-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-764-870-7
; Sequence 7, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletterick, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Kushner, Peter J
; APPLICANT: Apriletti, James W
; APPLICANT: West, Brian
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
; TITLE OF INVENTION: Binding Domains
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,540
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,543
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,606
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Jackie N
; REGISTRATION NUMBER: 35,966
; REFERENCE/DOCKET NUMBER: UCAL 246/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid

; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US 08-764-872 7

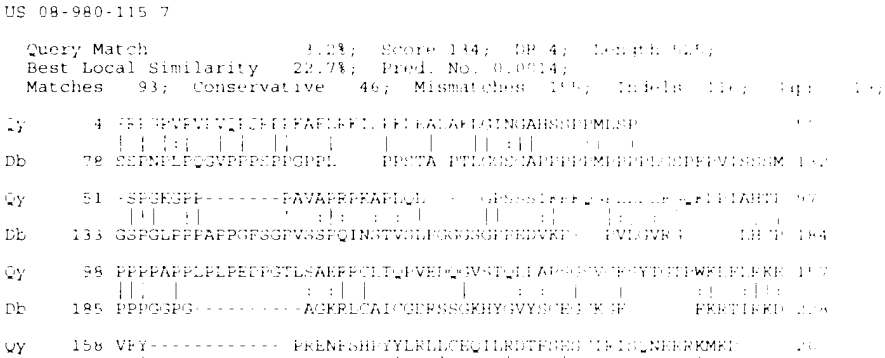
Query Match 3.2%; Score 134; DB 4; Length 525;
Best Local Similarity 22.7%; Prod. No. 0.0014;
Matches 93; Conservative 46; Mismatches 155; Indels 116; Gaps 19;



RESULT 2
US-08-980-115-7
; Sequence 7, Application US/08980115
; Patent No. 6276622
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S.
; APPLICANT: Barton, John D.
; APPLICANT: Fletterick, Robert J.
; APPLICANT: Wagner, Richard L.
; APPLICANT: Kushner, Peter J.
; APPLICANT: Apriletti, James W.
; APPLICANT: West, Brian L.
; APPLICANT: Shiau, Andrew K.
; TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
; FILE REFERENCE: UCAL 246/02US
; CURRENT APPLICATION NUMBER: US/08/980,115
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 08/764,872
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 60/008,606
; EARLIER FILING DATE: 1995-12-14
; EARLIER APPLICATION NUMBER: 60/008,543
; EARLIER FILING DATE: 1995-12-13
; EARLIER APPLICATION NUMBER: 60/008,540
; EARLIER FILING DATE: 1995-12-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (292)...(523)
; OTHER INFORMATION: minimal ligand binding domain

US 08-980-115 7

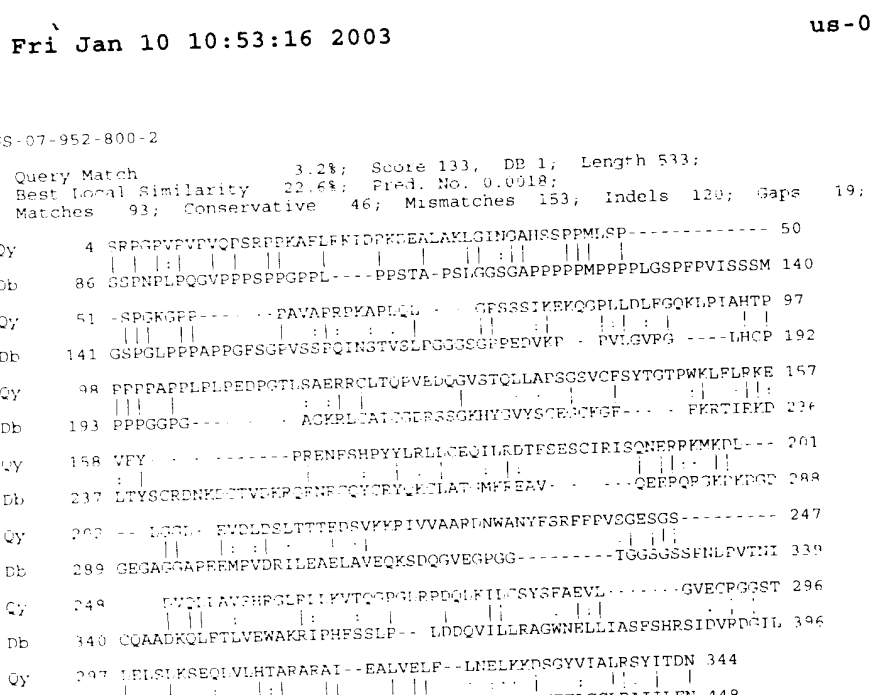
Query Match 3.2%; Score 134; DB 4; Length 525;
Best Local Similarity 22.7%; Prod. No. 0.0014;
Matches 93; Conservative 46; Mismatches 155; Indels 116; Gaps 19;



RESULT 3
US-07-952-800-2
; Sequence 2, Application US/07952800
; Patent No. 5403925
; GENERAL INFORMATION:
; APPLICANT: OZATO, KEIKO
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR H
; TITLE OF INVENTION: RECEPTOR SUPERFAMILY AND A
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND
; STREET: 1 MARKET PLAZA, STEUART TOWER, SUITE
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,800
; FILING DATE: 19920928
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-21-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein

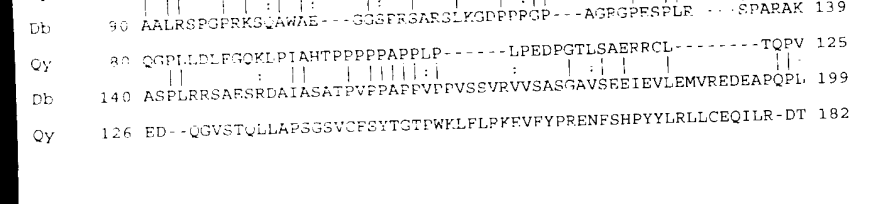
US-07-952-800-2

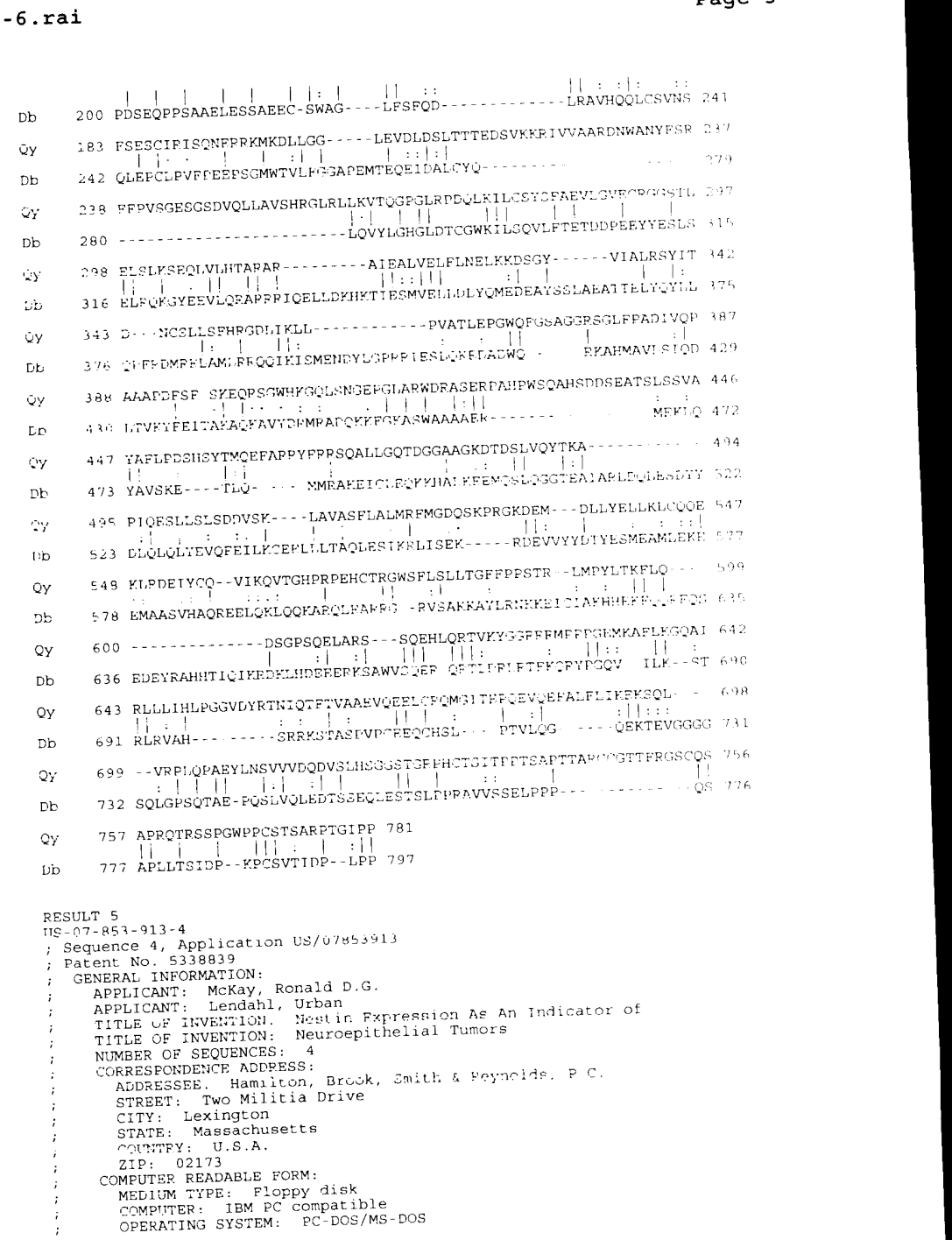
Query Match 3.2%; Score 133, DB 1; Length 533;
Best Local Similarity 22.6%; Pred. No. 0.0018;
Matches 93; Conservative 46; Mismatches 153; Indels 120; Gaps 19;



RESULT 4
US-09-311-236-2
; Sequence 2, Application US/09311236
; Patent No. 6376215
; GENERAL INFORMATION:
; APPLICANT: LaThangue, Nicholas B
; APPLICANT: Shikama, No 6376215iko
; APPLICANT: The University Court of the University of Glasgow
; TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
; TITLE OF INVENTION: JMY and uses thereof
; FILE REFERENCE: AHB/LP5731500
; CURRENT APPLICATION NUMBER: US/09/311,236
; CURRENT FILING DATE: 1999-05-13
; EARLIER APPLICATION NUMBER: GB 9722238.4
; EARLIER FILING DATE: 1997-10-21
; EARLIER APPLICATION NUMBER: GB 9818235.5
; EARLIER FILING DATE: 1998-08-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-311-236-2

Query Match 3.2%; Score 130.5; DB 4; Length 983;
Best Local Similarity 20.7%; Pred. No. 0.0078;
Matches 179; Conservative 109; Mismatches 316; Indels 261; Gaps 42;





RESULT 5
US-07-853-913-4
; Sequence 4, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/xxx,xxx
;       FILING DATE:  19920319
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/660,412
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/603,803
;       FILING DATE:  25 OCT 1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/201,762
;       FILING DATE:  02-JUN-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/180,548
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Granahan, Patricia
;       REGISTRATION NUMBER:  32,227
;       REFERENCE/DOCKET NUMBER:  MIT-4641AAAA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-861-6240
;       TELEFAX:  617-861 9540
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1618 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-853-913-4

  Query Match              3.0%;  Score 124;  DB 1;  Length 1618;
  Best Local Similarity   20.6%;  Pred. No. 0.866;
  Matches  163;  Conservative  105;  Mismatches  266;  Indels  256;  Gaps   41;

Qy        9 PVPVPVQPSRPPPAFLPPI                DPPRDEALARLQI                  38
            |||.|.:.:.|.:|..|:                |.|.:..:.|                    
Db      856 PSPLPAILEIPVPAPLRNVEPLQARTPTLA    STP  IPPTP QAPSPAVDAEI 904

Qy       46 P   APLQVGPSSSIKEKQGPLLDLFGVKLPIAHTPPP--------PPAPPLPLPEDPGT 113
            :   ''          :.:.:.|.      :.|.        | |         ||  || ||:||
Db      905 SA.DAPL      -SLDQTQG------GRR--  QAPEPLRAEAPVAIPASVLPGPEEPGG 948

Qy      114 PVATPPVTVQPVHTVQGVSTQKIAP             GSVVGFSY- - -TGTFWKFP -  163
            .   .  | |:|          |:|             ||  ||          |  |:|
Db      949 LKLRAATVQRP-PDHASLAPPLSPDHSSIPAPDGESGGSPVPSIGPGPGPGQIWGLVFKE 007

Qy      164       - --- LREEVT               YPRENFCHFYYLRLLIEQILRDTFSEGCIR 199
                    :::|::                 | |  :    |: | |:|      :  ::
Db      008 TAIPGYVVPGLQPIWPPPIGNPPPIAPVGVPIFPPT-   IPGIGPPI -----QPSLR 057

Qy      200 IVLNF      PPKMKDLIGNLPVDLDVLTTPDGVPPFIVVAAPLDWAHYFPPPPVGDE 244
            : :        |: ::    || ||::| . |   |. |                    |
Db      058 TLENLSHETLEPPNQFTPPSLPPHIPTLPSIPPPNPPAI- . ---  ------- -PGI 099

Qy      245 SASDVLLLAVSHRGLELLAVTLQPGLPPLLKILDLSYSPAEVLGVECPGGSTLELSLKSE 304
            ::      |: : || |   :     |                  ||:      |
Db      100 PPPPP-    PPPPPPPIPPT  -PPPPP- . .   .   .   .      TLPSLQPPN 681

Qy      305 QIVLHTRAAPAIFALVPLPLNPLPPPGYYIALPPYIPDNCSLLCPHPTLLPI--LPVA 362
            | ::      :::|  :| ||                  |:|  |:|   : :|  |  |
Db      682 QPLM     KGLEDNLETPLPP               GTPNQELVSSLQENLESLTALEKE 672

Qy      363 TLPPPWQPPPAVGPPSLPPAGIVQPAAAAPCFSFSFFQPSQWHFQGLSNCPP---GLAPWQR 420
            :: :       | |  : :     |  |  :  :   | ||    |    |:
Db      673 NQPP-PPPPPYPTPPVIPP  -TPPPNIPPIPCPIPTPNPPAIP-LEPPLNLPLPTLPPPQA 732

Qy      421 APPPAHPWQAHGGPAPSLPVAYAPLPGSPPYPMLPPAPPYPPPSLALPGLTLPGA 480
            :  |      :: |       || |     |  :    |::    :   :| :: ||: |
Db      733 SIVPPLP  TPNHN-     SLPS     PPPQPQPPLPPLPPPTQPPPPS-DGPQD--- 772

Qy      481 AGKVTMSLVLYTPAPIQ   -PSPISLSPDVSPLAVASPLALMPFMGPQSPPPGKDEMDL 536
                    : |  ' :    | ' ||  ::::         :: : ::|    |
```

```
Db      773  -       QMLPPPPPPVPCPPGPSLPVPIAPPLPNPN PPLPSLPPPSVPEAVPSLPTP 826

Qy      537 IYPLLPLCQQPKLPLPIPSQVIPVPVPPPIPTVPTPLPTPLMPTLTP 583
            : | ||   || |       : :|                          :   :
Db      824 ILPSLPSAGQPNL--- - PTLPS                    PPTVAPLWTPP 865

Qy      584 PLQQQGPSQPLARQQQPRLQKTVPPQGK  PPMPIPIMPPAPLPLPIAPPLLPLPPP 641
              : ||  ::    |||:   ||    |   | :  |,       :::    :| : | ::
Db      866 PINP-PPNP    -PPPPNPNPPLPVPPPPPPPIQTPP      PPPP   PPIPPPPMP 909

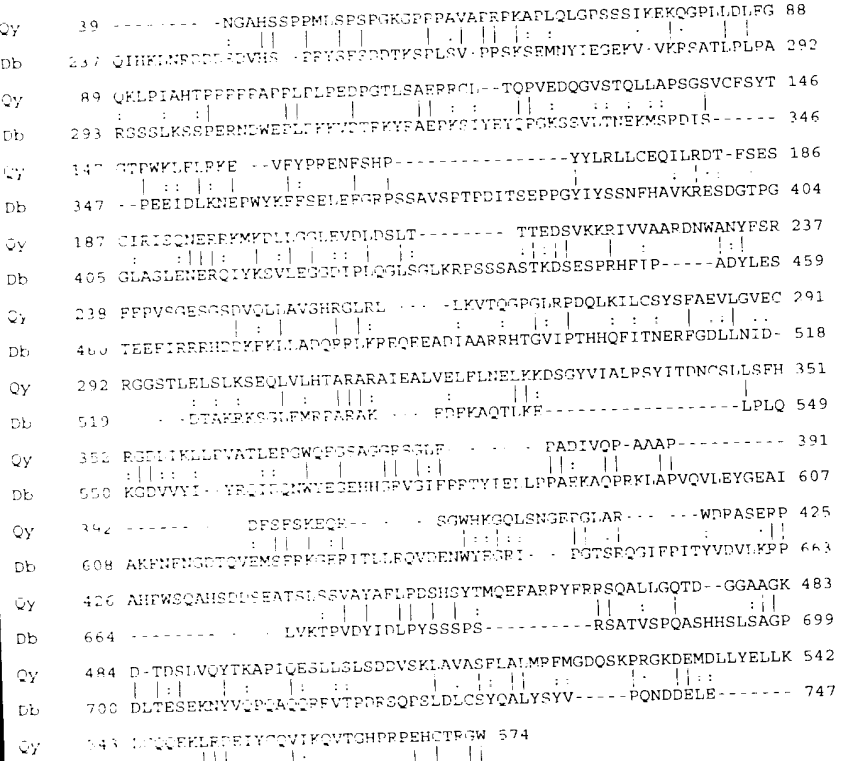
Qy      642 YKTNIQTPTVAAEVQPELPQRQMGITP   PQPVQPPALPLIPPPPPLVPPLQLPAPVINGV 701
                  | ::  | :|  | :    |     | ||     | :' :| :       : ::
Db      910     PGPPPIPPGVPPPPSPNLPPPPPNLPLPVPPPPPPPPPPPPPPPPPPPPPPPPPPP 969

Qy      702 VVPPPVPLHS 721
              ||:::  |
Db      970 EKDPPELAQES 965


RESULT 6
US-08-630-915A-30
; Sequence 30, Application US/08630915A
; Patent No. 6309820
;  GENERAL INFORMATION:
;    APPLICANT:  SPARKS, Andrew B.
;    APPLICANT:  HOFFMAN, No. 6309820h
;    APPLICANT:  KAY, Brian K.
;    APPLICANT:  FOWLKES, Dana M.
;    APPLICANT:  McCONNELL, Stephen J.
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING A FUNCTIONAL
;    TITLE OF INVENTION:  DOMAIN OF INTEREST AND METHODS FOR IDENTIFYING AND
;    TITLE OF INVENTION:  USING SAME
;    NUMBER OF SEQUENCES:  227
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,915A
;      FILING DATE:  03-APR-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  1101-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  788 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-630-915A-30

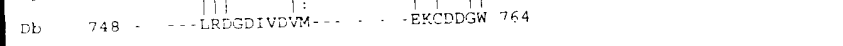
  Query Match              2.9%;  Score 119.5;  DB 4;  Length 788;
  Best Local Similarity   18.2%;  Pred. No. 0.095;
  Matches  126;  Conservative   96;  Mismatches  241;  Indels  229;  Gaps   47;

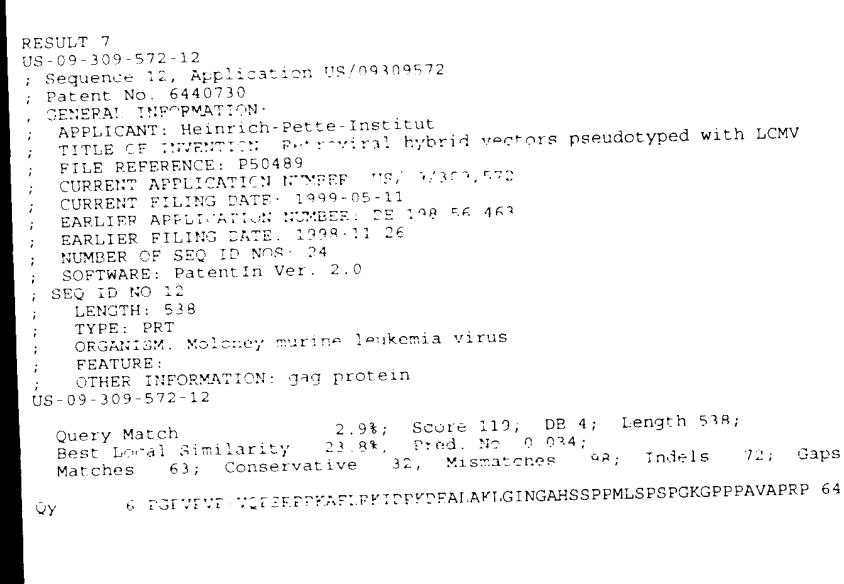
Qy        8 PVPVPVQPSRPPPAFLPPI                DPPRDEALARLQI                  38
            |||.|.:.:.|.:|..|:                |.|.:..:.|                    
Db      177 PVPTPARLTSPGRASERRARDASRRVVRSAQPLSPVSTLEVTILLINTRSPSIWYPTMPF 236
```

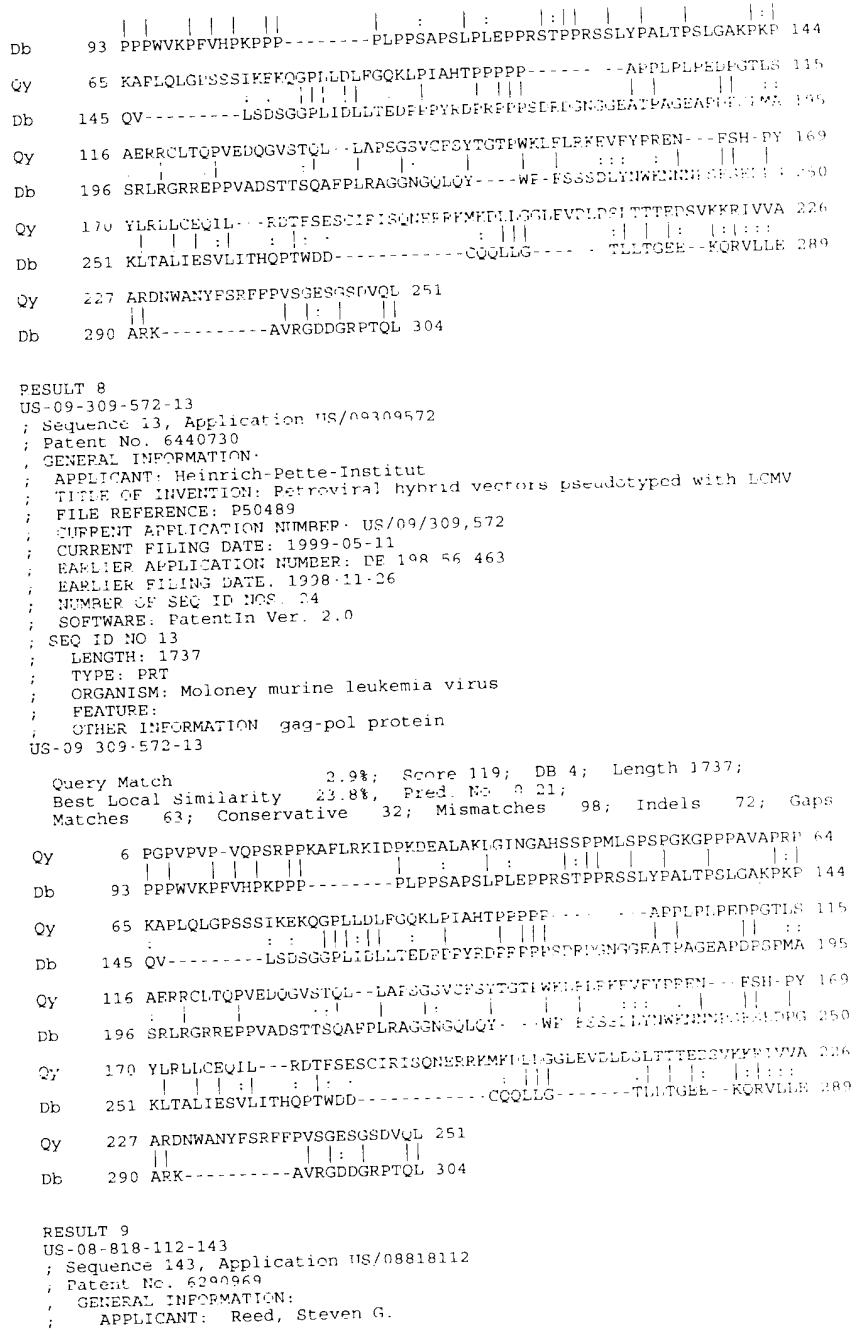


RESULT 7
US-09-309-572-12
; Sequence 12, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: gag protein
US-09-309-572-12

Query Match 2.9%; Score 119; DB 4; Length 538;
Best Local Similarity 23.8%; Pred. No. 0.034;
Matches 63; Conservative 32; Mismatches 98; Indels 72; Gaps 14;







PESULT 8
US-09-309-572-13
; Sequence 13, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1737
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: gag-pol protein
US-09-309-572-13

Query Match 2.9%; Score 119; DB 4; Length 1737;
Best Local Similarity 23.8%; Pred. No. 0.21;
Matches 63; Conservative 32; Mismatches 98; Indels 72; Gaps 14;

RESULT 9
US-08-818-112-143
; Sequence 143, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

```
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonio
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND DIAGNOSIS OF TUBERCULOSIS
;    NUMBER OF SEQUENCES:  153
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,111
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.41106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  138:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  174 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-818-111-138

  Query Match                 2.9%;  Score 118.5;  DB 4;  Length 174;
  Best Local Similarity   28.2%;  Pred. No. 0.0064;
  Matches   50;  Conservative    9,  Mismatches   47,  Indels   71,  Gaps    9;

Qy       5 RPGPVPVPVQPSRP         PPAPLPRPIPPPPPAPAPPGINGAHSSPPMLSPPPPGPGP 56
           ||   |||  |  |         |:| |                 || |: :|   |
Db      18 RPPVPPVPPPLPPSPPSPPTGWVPRALL           ----------PPWLAGTPPAPP 56

Qy      57 PPAVAP-RPPAPLQLGPSSSIPEPQGPILPLPGQFLPIAHTP----------------P 98
           | :||  | |||             ||  |    || :| |                 |
Db      57 VPPMAPLPPAAPLP          PLPPL--PPLPTSHPPPPPAPPAPPAPPACPPVP 102

Qy      99 PPPAPPLPLPELPGTLSAERRC-------LTQPVEDLVSTQLLAPSSVIPSYTS 147
           |||||| |  |  | |:  |           |  |    |  |  : :  |  |
Db     103 VPPAPPLP-PSPPTELPADAACTPAPPAPPLAPPSPPAG-SAAIPALTGATSASTLG 157


RESULT 10
US-08-818-111-138
; Sequence 138, Application US/08818111
; Patent No. 6338852
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos Neto, Antonio
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES:  148
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-056-556-143

Query Match 2.9%; Score 118.5; DB 4; Length 174;
Best Local Similarity 28.2%; Pred. No. 0.0064;
Matches 50; Conservative 9; Mismatches 47; Indels 71; Gaps 9;

Qy 5 RPGPVPVPVQPSRP------- PKAPLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGP 56
Db 18 RPPVPPVPPLPPSPPSPPTGWVPRALL--------------------PPWLAGTPPAPP 56

Qy 57 PPAVAP-RPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTP----------------P 98
Db 57 VPPMAPLPPAAPLP------------PLPPL--PPLPTSHPPRPPAPPAPPAPPACPFVP 102

Qy 99 PPPAPPLPLPEDPGTLSAERRC--------LTQPVEDQGVSTQLLAPSGSVCPSYTG 147
Db 103 VPPAPPLP-PSPPTELPADAACPPAPPAPPLAPPSPPAG-SAAIRALTGATSASTLG 157

RESULT 12
US-09-072-596-138
; Sequence 138, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A W
; APPLICANT: Dillon, Davin C
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-072-596-138

Query Match 2.9%; Score 118.5; DB 4; Length 174;
Best Local Similarity 28.2%; Pred. No. 0.0064;
Matches 50; Conservative 9; Mismatches 47; Indels 71; Gaps 9;

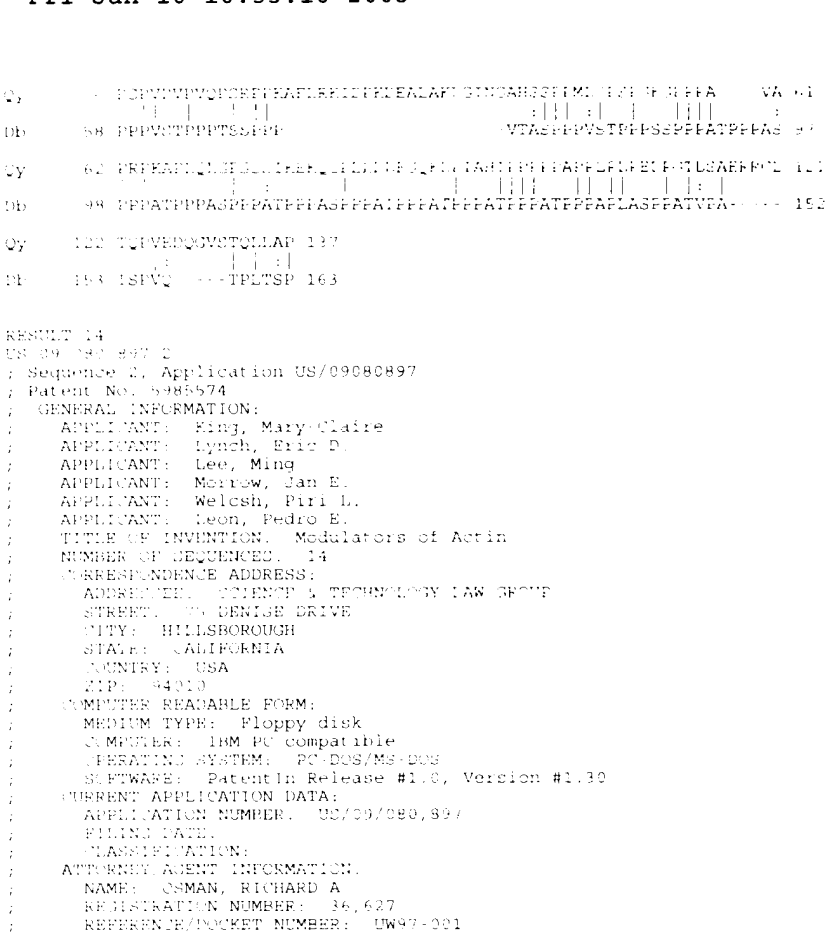
Qy 5 RPGPVPVPVQPSRP------- PKAPLRKIDPKDEALAKLGINGAHSSPPMLSPSPGKGP 56
Db 18 RPPVPPVPPLPPSPPSPPTGWVPRALL--------------------PPWLAGTPPAPP 56

Qy 57 PPAVAP-RPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTP----------------P 98
Db 57 VPPMAPLPPAAPLP------------PLPPL--PPLPTSHPPRPPAPPAPPAPPACPFVP 102

Qy 99 PPPAPPLPLPEDPGTLSAERRC--------LTQPVEDQGVSTQLLAPSGSVCPSYTG 147
Db 103 VPPAPPLP-PSPPTELPADAACPPAPPAPPLAPPSPPAG-SAAIRALTGATSASTLG 157

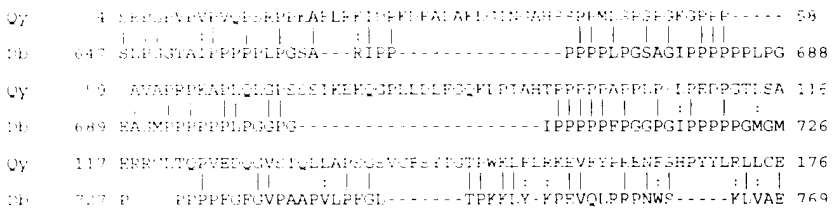
RESULT 13
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-217-327-4

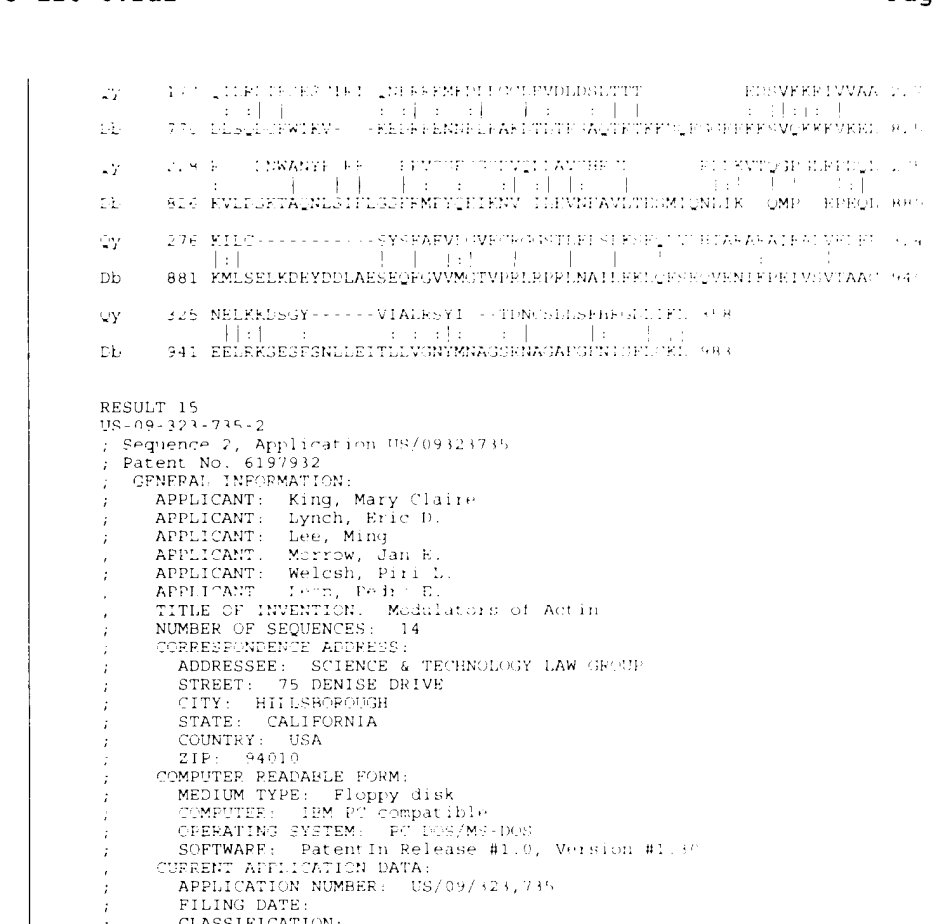
Query Match 2.9%; Score 118; DB 1; Length 214;
Best Local Similarity 27.9%; Pred. No. 0.0098;
Matches 38; Conservative 7; Mismatches 57; Indels 34; Gaps 4;



RESULT 14
US-09-080-897-2
; Sequence 2, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-080-897-2

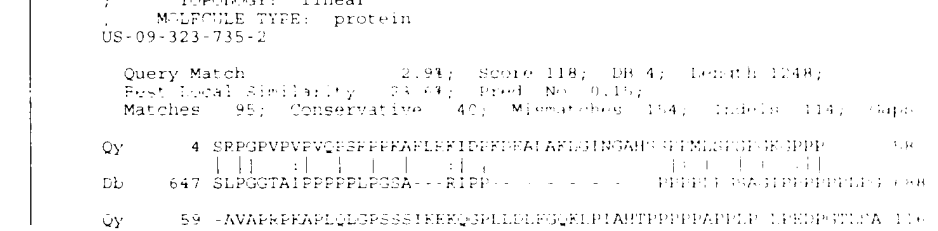
Query Match 2.9%; Score 118; DB 2; Length 1248;
Best Local Similarity 23.6%; Pred. No. 0.15;
Matches 95; Conservative 40; Mismatches 154; Indels 114; Gaps 20;

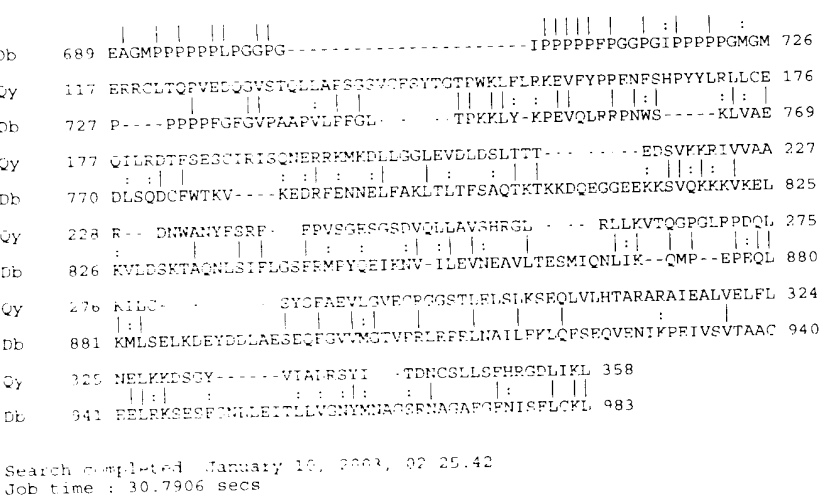




RESULT 15
US-09-323-735-2
; Sequence 2, Application US/09323735
; Patent No. 6197932
; GENERAL INFORMATION:
; APPLICANT: King, Mary Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/323,735
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/080,897
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-323-735-2

Query Match 2.9%; Score 118; DB 4; Length 1248;
Best Local Similarity 23.6%; Pred. No. 0.15;
Matches 95; Conservative 40; Mismatches 154; Indels 114; Gaps 20;





Search completed January 10, 2003, 09:25:42
Job time : 30.7906 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 02:23:04 ; Search time 60.7922 Seconds
(without alignments)
250.842 Million cell updates/sec

Title: US-09-803-126-6
Perfect score: 4129
Sequence: 1 MYQSRPGPVPVPVQPSRPPKAPLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 786

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4129	100.0	786	10	US-09-803-126-6	Sequence 6, Appli
2	2217.5	53.7	437	10	US-09-803-126-4	Sequence 4, Appli
3	2120	51.3	405	10	US-09-803-126-13	Sequence 13, Appl
4	1655.5	40.1	333	9	US-09-764-868-758	Sequence 758, App
5	1537	37.2	407	10	US-09-803-126-12	Sequence 12, Appl
6	1537	37.2	1337	10	US-09-803-126-1	Sequence 1, Appli
7	306	7.4	56	10	US-09-864-761-40685	Sequence 40685, A
8	153.5	3.7	1031	9	US-09-764-868-647	Sequence 647, App
9	143	3.5	559	10	US-09-858-156A-2	Sequence 2, Appli
10	131.5	3.2	1134	9	US-10-001-873-50	Sequence 50, Appl
11	130.5	3.2	983	10	US-09-912-917-2	Sequence 2, Appli
12	129	3.1	446	10	US-09-853-386-69	Sequence 69, Appl
13	129	3.1	643	10	US-09-853-386-70	Sequence 70, Appl
14	128.5	3.1	802	10	US-09-823-240-2	Sequence 2, Appli
15	124.5	3.0	507	9	US-10-078-547-24	Sequence 24, Appl
16	124.5	3.0	625	10	US-09-853-386-63	Sequence 63, Appl
17	124.5	3.0	626	10	US-09-853-386-64	Sequence 64, Appl
18	124.5	3.0	626	10	US-09-853-386-65	Sequence 65, Appl
19	124.5	3.0	626	10	US-09-853-386-96	Sequence 96, Appl
20	124.5	3.0	731	9	US-10-086-464-17	Sequence 17, Appl
21	124	3.0	21	10	US-09-803-126-8	Sequence 8, Appli
22	124	3.0	1618	9	US-09-963-875-1	Sequence 1, Appli
23	123	3.0	684	10	US-09-823-240-9	Sequence 9, Appli
24	121	2.9	647	10	US-09-825-299-1002	Sequence 1002, Ap
25	121	2.9	3256	10	US-09-919-172-98	Sequence 98, Appl
26	120.5	2.9	1567	10	US-09-835-232-2	Sequence 2, Appli
27	120	2.9	503	9	US-10-078-547-2	Sequence 700, App
28	119.5	2.9	200	10	US-09-825-297-700	Sequence 80, Appl
29	119.5	2.9	788	10	US-09-879-957-80	Sequence 43, Appl
30	119	2.9	463	9	US-10-029-180-80	Sequence 43, Appl
31	118	2.9	538	9	US-09-976-740-43	Sequence 43, Appl
32	118	2.9	538	12	US-10-023-529-43	Sequence 9, Appli
33	118	2.9	538	12	US-10-023-523-42	Sequence 40290, A
34	117	2.8	538	10	US-09-827-922-9	Sequence 2, Appli
35	116	2.8	116	10	US-09-864-761-40290	Sequence 4, Appli
36	116	2.8	659	9	US-10-144-621-2	Sequence 2, Appli
37	116	2.8	726	10	US-09-770-689A-4	Sequence 2, Appli
38	116	2.8	803	10	US-09-770-689A-2	Sequence 6, Appli
39	116	2.8	881	10	US-09-816-860A-2	Sequence 34988, A
40	116	2.8	1017	9	US-09-965-830-6	Sequence 36844, A
41	115.5	2.8	171	10	US-09-864-761-34988	Sequence 47, Appl
42	115.5	2.8	171	10	US-09-864-761-36844	Sequence 47, Appl
43	114	2.8	550	9	US-09-976-740-47	Sequence 47, Appl
44	114	2.8	550	12	US-10-023-529-47	
45	114	2.8	550	12	US-10-023-523-47	

ALIGNMENTS

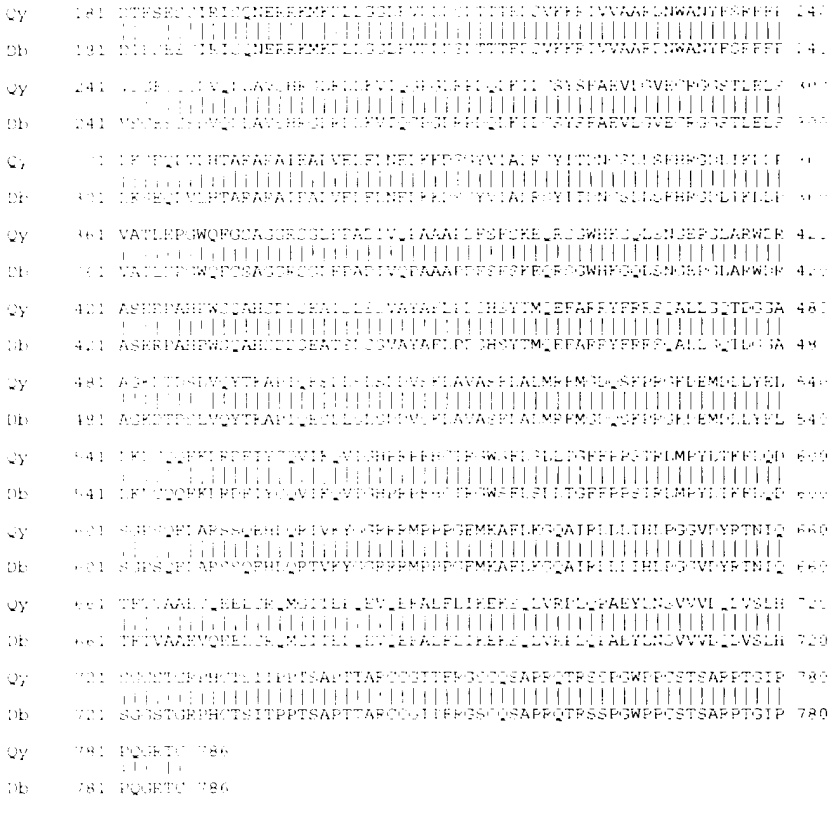
RESULT 1
US-09-803-126-6
; Sequence 6, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 2 (hMRP2)
US-09-803-126-6

Query Match 100.0%; Score 4129; DB 10; Length 786;
Best Local Similarity 100.0%; Pred. No. 3.5e-299;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MYQSRPGPVPVPVQPSRPPKAPLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MYQSRPGPVPVPVQPSRPPKAPLPKIDPKDEALAKLGINGAHSSPPMLSPSPGKGPPPAV 60

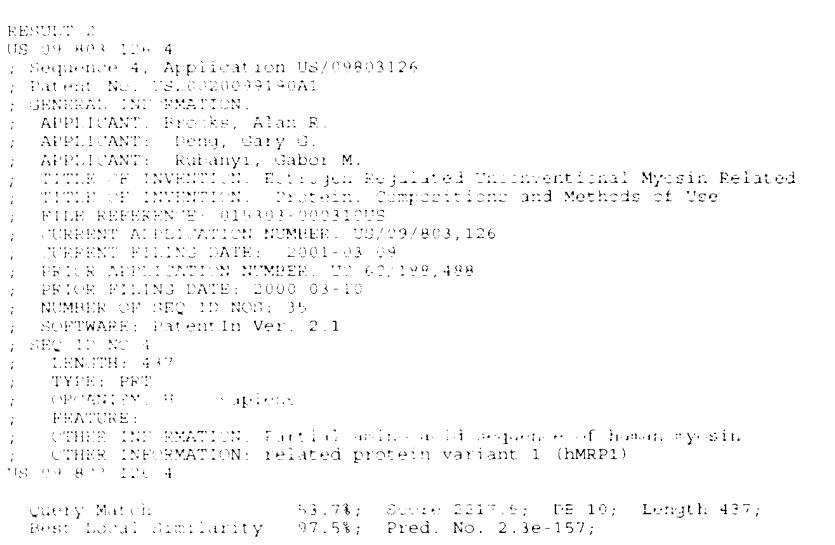
Qy      61 APRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 APRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPEDPGTLSAERRC 120

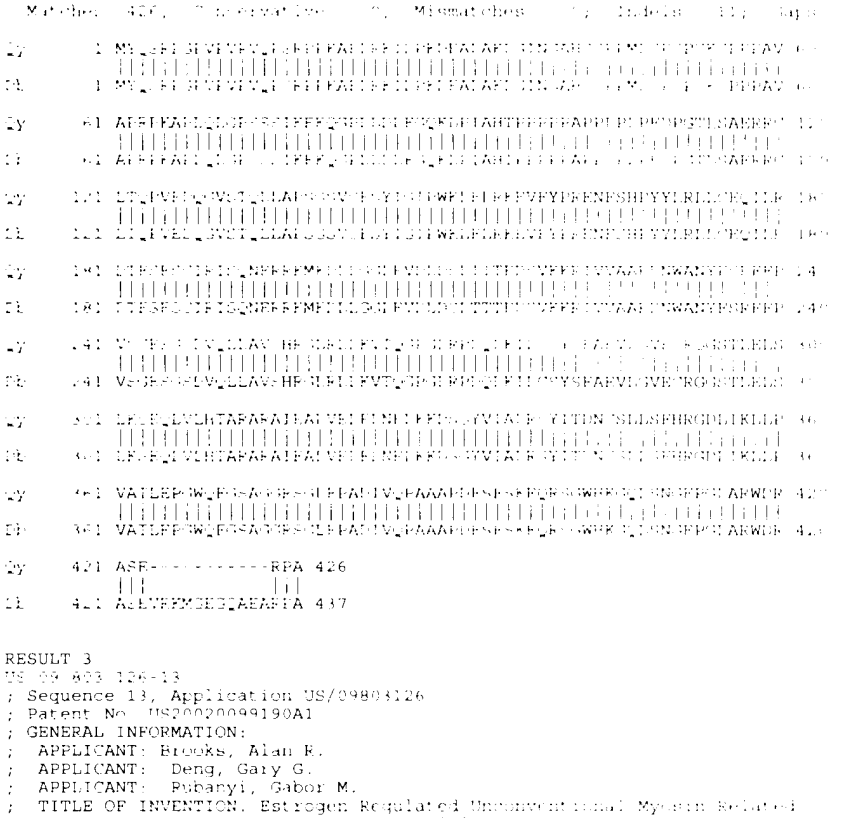
Qy     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLPLPFEVFYPRLNFSHFYYLRLLCEQILR 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLPLPFEVFYPRLNFSHFYYLRLLCEQILR 180
```



RESULT 2
US-09-803-126-4
; Sequence 4, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 1 (hMRP1)
US-09-803-126-4

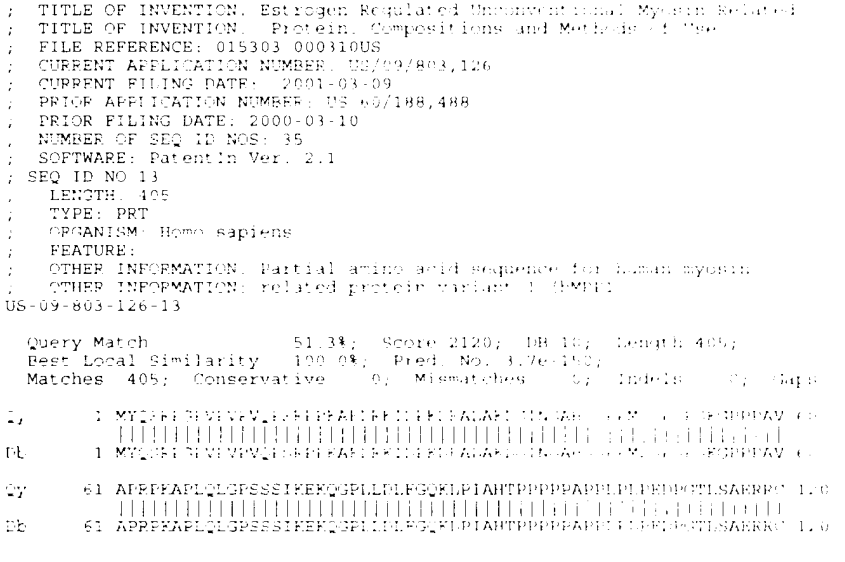
Query Match 53.7%; Score 2217.5; DB 10; Length 437;
Best Local Similarity 97.5%; Pred. No. 2.3e-157;





RESULT 3
US-09-803-126-13
; Sequence 13, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence for human myosin
; OTHER INFORMATION: related protein variant 1 (hMPF1)
US-09-803-126-13

Query Match 51.3%; Score 2120; DB 10; Length 405;
Best Local Similarity 100.0%; Pred. No. 3.7e-150;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
Qy     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLPKEVFYPRENFSHPYYLRLLCEQILR 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLPKEVFYPRENFSHPYYLRLLCEQILR 180

Qy     181 DTFSESCIRISQNEPPPMKDLLGGLEVDLDSLTTTEDSVKKRIVVAARDNWANYFSRFFP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DTFSESCIRISQNEPPPMKDLLGGLEVDLDSLTTTEDSVKKRIVVAARDNWANYFSRFFP 240

Qy     241 VSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECPGGSTLELS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECPGGSTLELS 300

Qy     301 LKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIKLLP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 LKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIKLLP 360

Qy     361 VATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQPSGWHK 405
           |||||||||||||||||||||||||||||||||||||||||||||
Db     361 VATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQPSGWHK 405


RESULT 4
US-09-764-868-758
; Sequence 758, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ32
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 758
;  LENGTH: 333
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-764-868-758

  Query Match             40.1%;  Score 1655.5;  DB 9;  Length 333;
  Best Local Similarity   96.7%;  Pred. No. 4.3e-116;
  Matches  320;  Conservative    0;  Mismatches    0;  Indels   11;  Gaps    1;

Qy     165 LPLPEDPGTLSAEPRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLPKEVFYPREN 164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 LPLPEDPGTLSAEPRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLPKEVFYPREN 60

Qy     165 FSHPYYLRLLCEQILRDTFSESCIRISQNEPPPMKDLLGGLEVDLDSLTTTEDSVKKRIV 224
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 FSHPYYLRLLCEQILRDTFSESCIRISQNEPPPMKDLLGGLEVDLDSLTTTEDSVKKRIV 120

Qy     225 VAARDNWANYFSRFFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFA 284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAARDNWANYFSRFFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFA 180

Qy     285 EVLGVECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDN 344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 EVLGVECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDN 240

Qy     345 CSLLSFHRGDLIKLLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWH 404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CSLLSFHRGDLIKLLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWH 300

Qy     405 KGQLSNGFFSLARWDRASE  - -- ----PPA 426
           ||||||||||||||||||||        |||
Db     301 KGQLSNGFFSLARWDRASEVPKMGEGQAEAPPA 333


RESULT 5
US-09-803-126-12
; Sequence 12, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
;  APPLICANT: Brooks, Alan R.
;  APPLICANT:  Deng, Gary G.
;  APPLICANT:  Rubanyi, Gabor M.
;  TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
;  TITLE OF INVENTION:  Protein: Compositions and Methods of Use
;  FILE REFERENCE: 015303-000310US
;  CURRENT APPLICATION NUMBER: US/09/803,126
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: US 60/188,488
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;  LENGTH: 407
;  TYPE: PRT
;  ORGANISM: Mus sp.
;  FEATURE:
;  OTHER INFORMATION: partial amino acid sequence for mouse myosin
;  OTHER INFORMATION: related protein (mMRP)
US-09-803-126-12

  Query Match             37.2%;  Score 1537;  DB 10;  Length 407;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-107;
  Matches  302;  Conservative   32;  Mismatches   70;  Indels    4;  Gaps    2;

Qy       1 MYQSRPGPVPVPVQPSRPPKAPLRKIDPKDEALAKLGINGAHSSPFMLSPSPGKGPPPAV 60
           |||||||| |||||:|| | | :| ||||||||||||||| |  |   ||: || ||||
Db       1 MYQSRPGPVAVPVQPTRPIKTPQKKNLPFDEALAKLGINGVH-LPLSTSPNQGKSPPAV 59

Qy      61 APPPKAPLQLGPSSSIKEKQGPLLDLFG---QKLPIAHTPPPPPAPPLPLPEDPGTLSAE 117
           |||||  :| || ||:|||||| ||||       | |  |||||| | ||  :| | | |
Db      60 VPRPKARPRLEPSLSIQEKQGPLRDLFGPCSPNPPTAPAPPPPPALPPPLSCEPKTPSVE 119

Qy     118 RRCLTQPVEDQGVSTQLLAPSGSVCFSYTGTPWKLFLRKEVFYPPENFSHPYYLRLLCEQ 177
           || |:!|: :||:|| |||||||||  |||||||||||||||||||| | ||||:|
Db     120 SHALTEPMEDKNISTKLLVPSGSVCFSYANAPWKLFLRKEVFYIPENFSHPYCLSLLCQQ 179

Qy     178 ILRDTFSESCIRISQNERRKMKDLLGGLEVDLDSLTTTEDSVKKRIVVAARDNWANYFSR 237
           |||||:||| ||||:|| ||| ||| ||| |::|    |||:||||||||||||||||||
Db     180 ILRDTFTESCTRISQDERHKMKGLLGDLEVSLETLDIVEDSIKKRIVVAARENWANYFSR 239

Qy     238 FFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLFPDQLKILCSYSFAEVLGVECPGGSTL 297
           |||||||||||||| ||||||||||||| |   |||| ||||:|||| |:||| |||
Db     240 IFPVSGESGSDVQLLGVSHRGLRLLKVTQSPSFHLDQLKTLCSYSYAEVLTVQCPGPSTL 299

Qy     298 ELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIK 357
           |||:|||:|||| ||:|:|:|||:||:|||||||||||||||||||||||||||||||||:
Db     300 ELSLKNEQLILHTAWARAIKAMVDLFLSELRKDSGYVIALRSYITDENSLLSFHPGDLIR 359

Qy     358 LLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHK 405
           |||| ||||||||||||||||| |:|||||||| |||  :|: | :
Db     360 LLPVTALEPGWQFGSAGGRSGLFPDDVVQPAAAPDLSFSLGKRNSWQR 407


RESULT 6
US-09-803-126-1
; Sequence 1, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
;  APPLICANT: Brooks, Alan R.
;  APPLICANT:  Deng, Gary G.
;  APPLICANT:  Rubanyi, Gabor M.
;  TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
;  TITLE OF INVENTION:  Protein: Compositions and Methods of Use
;  FILE REFERENCE: 015303-000310US
;  CURRENT APPLICATION NUMBER: US/09/803,126
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: US 60/188,488
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
```

; LENGTH: 1337
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: mouse myosin related protein (MRP)
US-09-803-126-2

Query Match 27.2%; Score 1035; DB 11; Length 1337;
Best Local Similarity 74.0%; Pred. No. 4.7e-108;
Matches 731; Conservative 30; Mismatches 71; Indels 4; Gaps 2;

Qy 1 MVQVRFQPVEVDQSEEPAAIRPFLEKLPAALAELRNDAHSFPMLFPSPGPGPEPAV 60
Db 914 MLSEIRVAVDVLSPKEPSFLPNSIFALAALSINDYR LRISRDFN SFSGPAV 973

Qy 61 AVDPAALSSFRSLRFLSRTLRP 82
Db 974 VSPPAAPPRIRPSPSIRSQPPQSPSRLPSPSASNFPTAPAPPPPAIPPPLSGPFPTPVP 1033

Qy 118 PSPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 177
Db 1034 PHALTPPNEPPLLCTRLLVCLSVPSTAGAIWELPLPPEVFTPPEENPCHPTCLSLLCQQ 1093

Qy 178 LLPNTPGPGGIRLCNFPPPPMTLLSTLPVSLDPLTTTPETCVPPPPIVVAAPPDWANVPGP 237
Db 1094 LLPGTPTPGTPRLCNFPPPPMPLLSTLPVSLETLPIVPPSIPPPPIVVAAPPDWANVPGP 1153

Qy 238 PPPVVPPQPPTVQLLAVPHPGLPPLPVTQSPTLPPPQLPPLPLCPSCPAEVLGVCTGGSTL 297
Db 1154 PPPVSGRGPPTQLPMPHPGLPPLPVTLPSPPHLLPPTLPTPSPIAPVLIVLPPPPPTL 1213

Qy 298 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 357
Db 1214 ELSLENEQLILHTAWAPAIPAWVTPLPLELPPLGTVIALPSYITDCNSLLSPHPGLIP 1273

Qy 358 LLPVTALPPPWQPSAGPRGRLPPAPIVQPAAAPDPSPPPEQPSGWHP 405
Db 1274 LLPVTALPPPWQPSAGPRGRLPPAPIVQPAAAPDPSPPPEQPSGWHP 1321

RESULT 7
US-09-864-761-40685
; Sequence 40685, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID SEQUENCES ENCODING PEP
; TITLE OF INVENTION: TISSUE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica X-2
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/180,310
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,452
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,364
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/744,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40685
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL 1.5
; OTHER INFORMATION: SWISSPROT HIT: Q90663, EVALUE 3.00e-08
; OTHER INFORMATION: EST_HUMAN HIT: AV644879.1, EVALUE 5.00e-28
US-09-864-761-40685

Query Match 7.4%; Score 306; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 1.9e-16;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 368 WQPGSAGGRSGLPPAPIVQPAAAPDPSPPPEQPSGWHP 405
Db 1 WQPGSAGGRSGLPPAPIVQPAAAPDPSPPPEQPSGWHP 38

RESULT 8
US-09-764-868-647
; Sequence 647, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 3110
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 647
; LENGTH: 1031
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (636)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-647

Query Match 7.2%; Score 295.5; DB 9; Length 1031;
Best Local Similarity 24.2%; Pred. No. 0.0026;
Matches 55; Conservative 40; Mismatches 99; Indels 37; Gaps 7;

Qy 498 IVQYIFAGIPPGLL 511

```
Db    505 ILRYTHHPLHSPLLPLPYGDINLNLLKDKGYTTLQDPAIKI----------FNSLQQLES 554

Qy    538 GKDEMDLDYELLKLCGD-EFLRDSIVCQVIKQVTGHPPPEHCTR--GWSFLSLLTGFFPP 586

Db    555 MGDPIPIIQGILQTGHDIFPLPDELYSQLIKQTNKVPHPGSVGNLYSWQILTCLSCTFLP 614

Qy    587 STRLMFYL--TKPLIDSGFSIFLARSGIEHLQRTVEYGGPEFMPPPGEMKAFLKGQAIR 643

Db    615 SRGILKYLKPHLFPIPEQPPGKEMEKYALPTYESLFKTKCREFVPSRDEIEALIHRQEM- 673

Qy    644 LLLIHLPGGVDYRTNIQTFTVAAEVQEELCRQMGITEPQEVQEFALF 690

Db    674 TSTVYCHGGGSCKITINSHTTAGEVVEKLIRGLAMEDSRNM--FALF 718


RESULT 9
US-09-858-155A-2
; Sequence 2, Application US/09858155A
; Patent No. US20020143143A1
; GENERAL INFORMATION:
;  APPLICANT: Mark, Robert
;  APPLICANT:  Young, Kathleen H
;  APPLICANT:  Wood, Andrew
;  TITLE OF INVENTION. PAPLO, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND
;  TITLE OF INVENTION:  USES RELATED THERETO
;  FILE REFERENCE: GNN-005
;  CURRENT APPLICATION NUMBER: US/09/858,155A
;  CURRENT FILING DATE:  2001-05-15
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 559
;   TYPE: PRT
;   ORGANISM. Homo sapiens
US-09-858-155A-2

  Query Match             3.5%;  Score 143;  DB 10;  Length 559;
  Best Local Similarity   24.5%;  Pred. No. 0.0066;
  Matches   67;  Conservative   36;  Mismatches   97;  Indels   74;  Gaps   15;

Qy      1 MYQGRP  -  GPVFVPVQPSPPPYAPLPKIDPKDEALAKLGINGAHSS--PPMLSPSPG 53

Db    303 LIENRPQSPATGRTPVFVSPTPPPP- -- PPLPSALSTSSLPASMTSTPPPPVPPPP- 356

Qy     54 KGPPPAVA------PRPKAPLQLGP----------SSSIKEKQGPLLDLFGQKLPIAHT- 96

Db    357  -PPPATALQAPAVPPPPAPLQIAPGVLHPAPPPIAPPLVQPSPPV----ARAAPVCETV 410

Qy     97 ---- -- -- -  PPPPPAPPLP -- LPEDPGTLSAEPPCLTQPVEDQGV-STQLLAPS 138

Db    411 PVHPLPQGEVQGLPPPPPPPPLPPPGIRPSSPVTVTA----LAHP--PSGLHPTPSTAPG 464

Qy    139 GSVCPSYTGTPWKLPLRKEVPYPPENPSHPYYLRLLCE--QILRDTPSESCIRISQNERR 196

Db    465 PHVPLMPPSPP--  - SQVIPASEPERHPSTLPVISDARSVLLEAIPKGIQLPKVPEQR 518

Qy    197 EMKDLLGGLEVDLDSLTTIEDSVKKFIVVAAPPN 230

Db    519 EQEAPHEPIENPVATI  -  -LSPPIAVEYSDS 546


RESULT 10
US-10-001-873-50
; Sequence 50, Application US/10001873
; Patent No. US20020160388A1
; GENERAL INFORMATION:
;  APPLICANT: Macina, Roberto
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Chen, Sui-Yu
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Turner, Leah
;  TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
;  FILE REFERENCE. DEX-0275
;  CURRENT APPLICATION NUMBER: US/10/001,873
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/252,055
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: 60/252,496
;  PRIOR FILING DATE: 2000-11-22
;  NUMBER OF SEQ ID NOS. 55
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
;   LENGTH: 1134
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-001-873-50

  Query Match             3.2%;  Score 131.5;  DB 9;  Length 1134;
  Best Local Similarity   34.3%;  Pred. No. 0.13;
  Matches   48;  Conservative    4;  Mismatches   65;  Indels   23;  Gaps    6;

Qy      4 SRPGPVFVPVQPS-- -- -- -  -PPPKAPLPKIDPKDEALAKLGIN---GAHSSPPML 48

Db    886 SAAAPAPVSTAPATKTSRCAKGGPLSPEKDAGEAYDPKDPKPKPKKGKEAGPGAGLPEERA 945

Qy     49 S- PSPGKGPPPAVAPRPKAPLQLGPSSSIKEKQGPLL--DLFGQKLPIAHTPPPPPAFP 104

Db    946 PALPSEAPAPPPPPPPPPHPPL---PPPPLPPPPLPLRLPPLPPPPLPREHPPPPPPLPP 1002

Qy    105 LPLPEDPGTLSAEPPCLTQP 124

Db   1003 LLPPPQTRTLPA-APTMPQP 1021


RESULT 11
US-09-912-917-2
; Sequence 2, Application US/09912917
; Patent No. US20020028487A1
; GENERAL INFORMATION:
;  APPLICANT: LaThangue, Nicholas B
;  APPLICANT:  Shikama, No. US20020028487A1ko
;  APPLICANT:  The University Court of the University of Glasgow
;  TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
;  TITLE OF INVENTION:  JMY and uses thereof
;  FILE REFERENCE. AHB/LP5731500
;  CURRENT APPLICATION NUMBER: US/09/912,917
;  CURRENT FILING DATE.  2001-07-25
;  PRIOR APPLICATION NUMBER: 09/311,236
;  PRIOR FILING DATE: 1999-05-13
;  PRIOR APPLICATION NUMBER. GB 9818235.5
;  PRIOR FILING DATE: 1998-08-20
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 983
;   TYPE: PRT
;   ORGANISM: Mus sp.
US-09-912-917-2

  Query Match             3.2%;  Score 130.5;  DB 10;  Length 983;
  Best Local Similarity   20.7%;  Pred. No. 0.13;
  Matches  179;  Conservative  109;  Mismatches  316;  Indels  261;  Gaps   42;

Qy     21 AFLRKIDP-KDEALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRPKAPLQLGPSSSIKEK 79

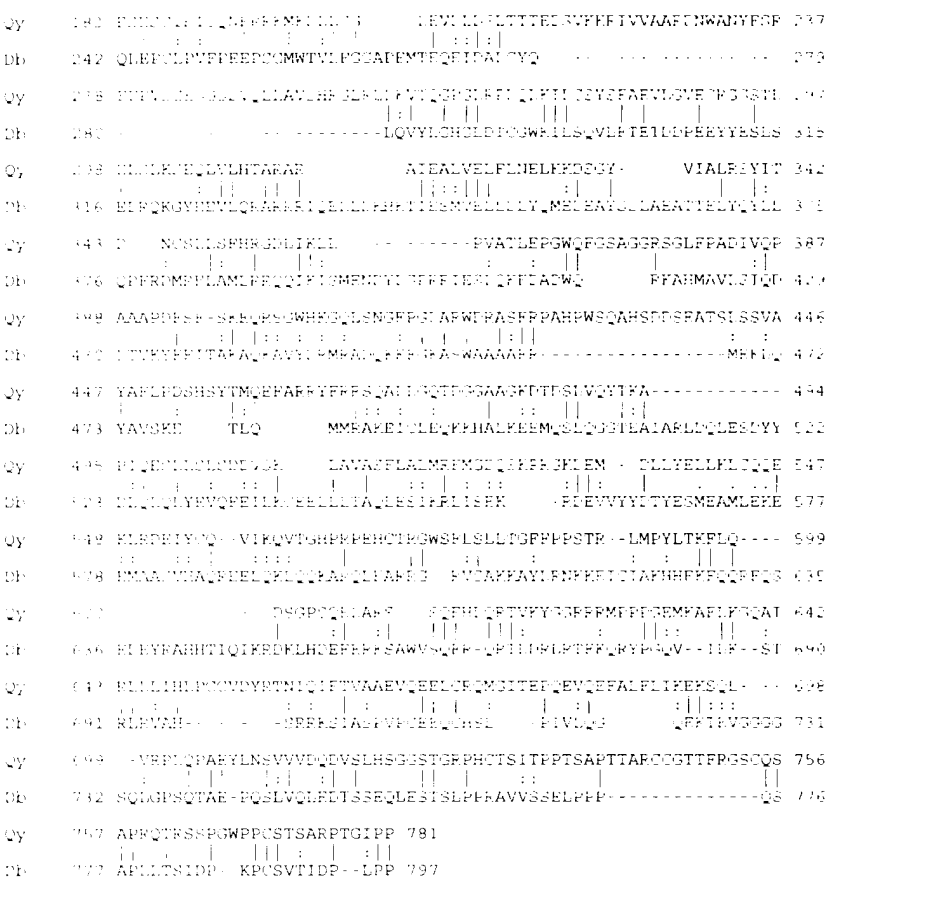
Db     90 AALRSPGPRKSQAWAE---GGSPRSAPSLKGDPPPGP---AGPGPPSPLR---  SPARAK 139

Qy     80 QGPLLDLFGQKLPIAHTPPPPPAPPLP------LPEDPGTLSAERRCL          TQPV 125

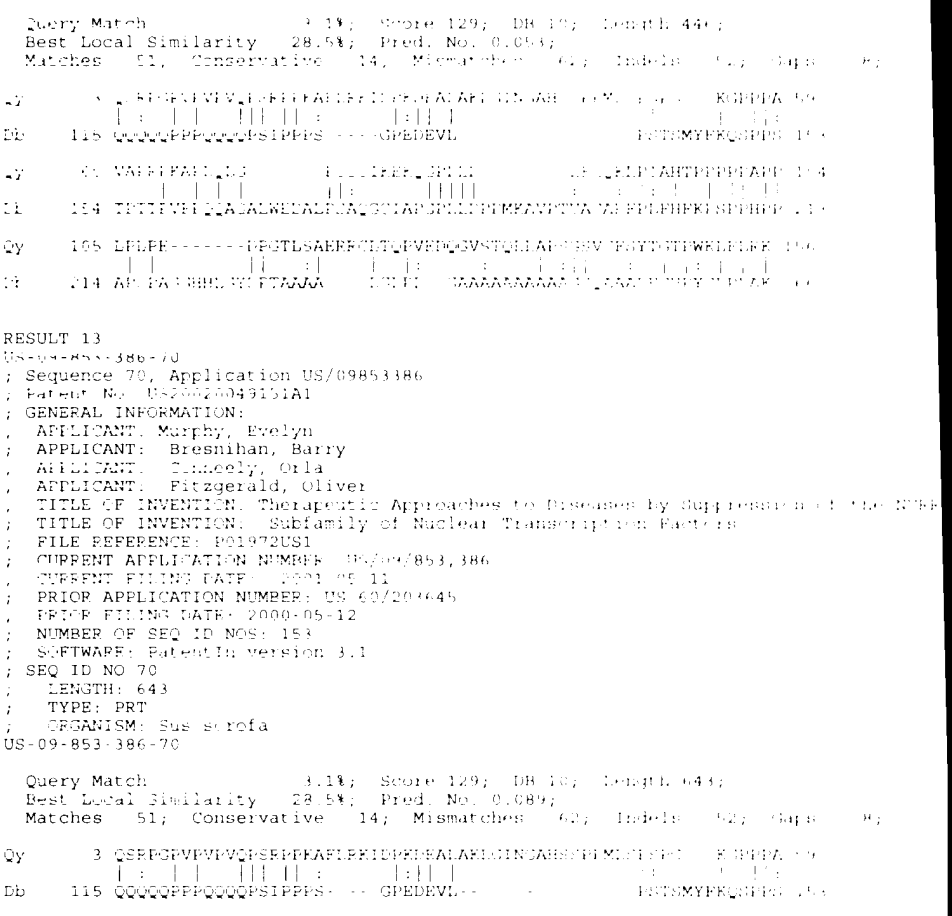
Db    140 ASPLPPSAESPDAIASATPVPPAPPVPPVSSVRVVSASGAVSEEIEVLEMVREDEAPQPL 199

Qy    126 ED--QGVSTQLLAPSGSVCPSYTGTPWKLPLRKEVPYPRENPSHPYYLRLLCEQILRDT 182

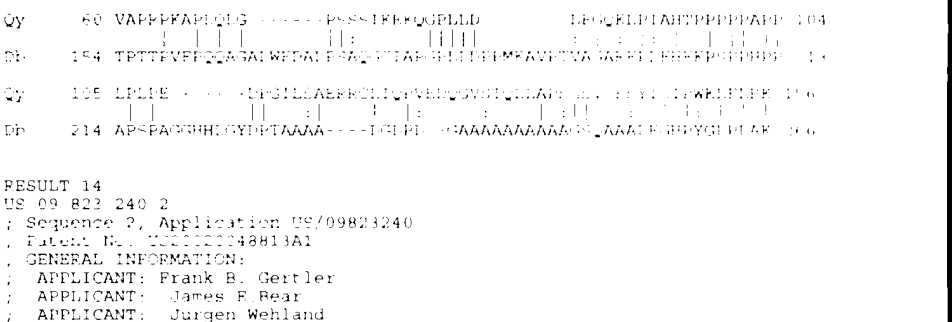
Db    200 PDSEQPPSAAELESSAEEC-SWAG----LFSFQD----------  LRAVHQQLCSVYS 241
```



RESULT 12
US-09-853-386-69
; Sequence 69, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Evelyn
; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NUPR
; TITLE OF INVENTION: Subfamily of Nuclear Transcription Factors
; FILE REFERENCE: P01972US1
; CURRENT APPLICATION NUMBER: US/09/853,386
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Sus scrofa
US-09-853-386-69



RESULT 13
US-09-853-386-70
; Sequence 70, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Evelyn
; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NUPR
; TITLE OF INVENTION: Subfamily of Nuclear Transcription Factors
; FILE REFERENCE: P01972US1
; CURRENT APPLICATION NUMBER: US/09/853,386
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Sus scrofa
US-09-853-386-70



RESULT 14
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureio
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30

```
;  PRIOR APPLICATION NUMBER: 60/194,564
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;    LENGTH: 802
;    TYPE: PRT
;    ORGANISM: Mus musculus
US-09-823-240-2

  Query Match             3.1%;  Score 128.5;  DB 10;  Length 802;
  Best Local Similarity   31.4%;  Pred. No. 0.13;
  Matches   37;  Conservative    4;  Mismatches   38;  Indels   39;  Gaps    4;

Qy       6 PGPVFVFVQPSRPPKAFLPKIDPKDPEALAKLGINGAHSSPPMLSPSPGKGPPPAVAPRPK 65
           | | | |  || |  |:   : |                ||   | |  ||||   | |
Db     543 PAPPPPPPLPSGP--AYASALPPP------------PGPPPPPPLPSTGPPPPPPPPPP 587

Qy      66 AFLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPAPPLPLPE-DPGTLSAERRCLT 122
           | |  |                 ||||||||||      |: | : | ||
Db     588 LPNQAPP----------------------PPPPPPAPPLPASGIFSGSTSEDNRPLT 622


RESULT 15
US-10-078-547-24
; Sequence 24, Application US/10078547
; Publication No. US20020199211A1
; GENERAL INFORMATION:
;  APPLICANT: Narayanaswamy Ramesh
;  APPLICANT:  Miguel A. de la Fuente
;  APPLICANT:  Ines M. Anton
;  APPLICANT:  Raif S. Geha
;  TITLE OF INVENTION: WIP, A WASP-Associated Protein
;  FILE REFERENCE: 1242.1022-005
;  CURRENT APPLICATION NUMBER: US/10/078,547
;  CURRENT FILING DATE:  2002-02-19
;  PRIOR APPLICATION NUMBER: 09/599,287
;  PRIOR FILING DATE: 2000-06-22
;  PRIOR APPLICATION NUMBER: PCT/US98/27501
;  PRIOR FILING DATE: 1998-12-22
;  PRIOR APPLICATION NUMBER: 60/101,457
;  PRIOR FILING DATE: 1998-09-23
;  PRIOR APPLICATION NUMBER: 60/068,533
;  PRIOR FILING DATE: 1997-12-23
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
;    LENGTH: 507
;    TYPE: PRT
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Translated WIP ORF No. US20020199211A1 3
US-10-078-547-24

  Query Match             3.0%;  Score 124.5;  DB 9;  Length 507;
  Best Local Similarity   25.1%;  Pred. No. 0.14;
  Matches   46;  Conservative   15;  Mismatches   45;  Indels   77;  Gaps    8;

Qy       3 QSRPGPVFVFVQPSRPPKAFLPKIDP-  ----------KDEA--LAKLGINGAHSSPP 46
           |:   |||    :|| | :  ||   |                ||   | :  :: : |:||
Db     291 QNNKPPVPSTPPPSAPHPPHLPPPPPSPPGPPPLPPSSSGNDETPRLPQRNLSLSSSTPP 350

Qy      47 MLSPSPGKG---PPPAVAPRPKAPLQLGPSSSIKEKQGPLLDLFGQKLPIAHTPPPPPA- 102
           :  ||||:    |||  : ||  |::  |  |     |||                |||||
Db     351 L--PSPGRSGPLPPPVPSERPPPPVRDPPGRS-----GPL------------PPPPPVS 390

Qy     103 ----------------------------PPLP--------LPEDPGTLSAERRCLTQPV 125
                                       ||||         |  | : |         |
Db     391 RNGSTSRALPATPQLPSRSGVPPPAGPPPPLPPDPPSAGAPPPPPPSTSIPNGFQDSPC 450

Qy     126 EDQ 128
           |||
Db     451 EDE 453
```

```
Search completed: January 10, 2003, 02:29:11
Job time : 65.7922 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 03:03:27 ; Search time 9191.4 Seconds
(without alignments)
11968.650 Million cell updates/sec

Title: US-09-803-126-7
Perfect score: 3780
Sequence. 1 cggcagcagcaggctgtgtt ... aaaaaaaaaaaaaaaaaaaa 3780

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched 2054640 seqs, 14551402978 residues

Total number of hits satisfying chosen parameters. 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3780	100.0	3780	6	AX253084	AX253084 Sequence
2	3352	88.7	4174	6	AX253082	AX253082 Sequence
3	3167	83.8	4208	9	AF418293	AF418293 Homo sapi
4	2763	73.1	4602	9	AF418292	AF418292 Homo sapi
5	1937	51.2	2347	9	BC027875	BC027875 Homo sapi
6	1856.4	49.1	1891	9	AK026339	AK026339 Homo sapi
7	1816.4	48.1	6004	9	AB058686	AB058686 Homo sapi
8	1808.4	47.8	2372	9	AK027846	AK027846 Homo sapi
9	1750.2	46.3	1767	9	HSM803107	AL713730 Homo sapi
10	1691.4	44.7	7925	9	AF418290	AF418290 Homo sapi
11	1490	39.4	6293	6	AX253079	AX253079 Sequence
12	1376.6	36.4	1998	9	AK096869	AK096869 Homo sapi
13	1135.6	30.0	7515	9	AF418286	AF418286 Homo sapi
14	1097.4	29.0	2099	9	AK025259	AK025259 Homo sapi
15	946.4	25.0	4375	6	AX253080	AX253080 Sequence
16	763.6	20.2	1547	10	BC024506	BC024506 Mus muscu
17	442	11.7	170056	2	AC087643	AC087643 Homo sapi
18	442	11.7	178515	2	AC015551	AC015551 Homo sapi
c 19	442	11.7	215434	2	AC019214	AC019214 Homo sapi
20	442	11.7	217584	2	AC087749	AC087749 Homo sapi
21	295.4	7.8	11876	9	AF144094	AF144094 Homo sapi
c 22	295.4	7.8	170056	2	AC087643	AC087643 Homo sapi
23	293	7.8	5709	9	AF418289	AF418289 Homo sapi
c 24	292	7.7	311	6	AX330132	AX330132 Sequence
c 25	292	7.7	311	6	AX332694	AX332694 Sequence
c 26	292	7.7	311	6	AX333379	AX333379 Sequence
c 27	292	7.7	311	6	AX408873	AX408873 Sequence
28	278.6	7.4	11769	10	AF144095	AF144095 Mus muscu
29	201.4	5.3	5351	10	AF053130	AF053130 Mus muscu
c 30	112.8	3.0	180956	2	AC123410	AC123410 Rattus no
31	112.8	3.0	200832	2	AC130970	AC130970 Rattus no
32	102.2	2.7	3110	9	HSM803101	AL713794 Homo sapi
33	101.4	2.7	147804	10	AL645647	AL645647 Mouse DNA
c 34	101.4	2.7	260310	2	AL645852	AL645852 Mus muscu
c 35	73.4	1.9	7218	6	I66494	I66494 Sequence 14
c 36	73.4	1.9	226714	2	AC079532	AC079532 Mus muscu
37	72.4	1.9	125020	9	AF429315	AF429315 Homo sapi
38	71.8	1.9	202370	2	AC102849	AC102849 Mus muscu
39	70.6	1.9	186400	2	AC095843	AC095843 Rattus no
c 40	70	1.9	125020	9	AF429315	AF429315 Homo sapi
41	69.2	1.8	207683	2	AC098712	AC098712 Mus muscu
c 42	69	1.8	123380	2	AC126533	AC126533 Rattus no
43	68.8	1.8	139838	2	AC128371	AC128371 Rattus no
44	68.4	1.8	126374	9	HS73J07	AL353999 Homo sapi
45	67.8	1.8	69070	2	AC099959	AC099959 Mus muscu

ALIGNMENTS

RESULT 1
AX253084
LOCUS AX253084 3780 bp DNA linear PAT 05-OCT 2001
DEFINITION Sequence 7 from Patent WO0168866.
ACCESSION AX253084
VERSION AX253084.1 GI:15986280
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3780)
AUTHORS Brooks,A.R., Deng,G.G. and Rubanyi,G.M.
TITLE Estrogen-regulated unconventional myosin related protein:
compositions and methods of use

```
  JOURNAL    Patent: WO 0168866-A 2 20-SEP-2001;
             SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..3780
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /note "Partial DNA sequence for human myosin related
                     protein variant 2 (hMRP2)"
BASE COUNT      838 a   1228 c   1107 g    607 t
ORIGIN

  Query Match             100.0%;  Score 3780;  DB 6;  Length 3780;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3780;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTAGCAGCAGGCTCGGGCCTCCGAGGCTGGCGTCCCAGGCCTCACCCTCAGCCGTCACC 60
Db       1 CGGCAGCAGCAGGCTCGGGCCTCCGAGGCTGGCGTCCCAGGCCTCACCCTCAGCCGTCACC 60

Qy      61 TTCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120
Db      61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120

Qy     121 GATGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
Db     121 GATGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180

Qy     181 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
Db     181 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240

Qy     241 ATGAAGCTCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCCCAGGAGGAGGAGGAGGAG 300
Db     241 ATGAAGCTCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCCCAGGAGGAGGAGGAGGAG 300

Qy     301 GAAGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360
Db     301 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360

Qy     361 CATCCTCCCCCCATCGTGAAGAAGCCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
Db     361 CCTCCTCCCCCCATCGTGAAGAAGCCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420

Qy     421 ATTGAGAGCTGAGCCAGCAAGGAGACAGGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480
Db     421 GCTGAGAGCTGAGCCAGCAAGGAGACAGGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480

Qy     481 TCAAGGGGACTTGTGGTGCGCAGTCAGACTCCAAGCCCAAGGGGCCACAACCCAGCAGGG 540
Db     481 TCAAGGGGACTTGTGGTGCGCAGTCAGACTCCAAGCCCAAGGGGCCACAACCCAGCAGGG 540

Qy     541 AAATTGGCAACATCATCGGCATGTACCAGAGCGGCGGGGCCGGTGCCTGTGCCGTGCC 600
Db     541 AAATTGGCAACATCATCGGCATGTACCAGAGCGGCGGGGCCGGTGCCTGTGCCGTGCC 600

Qy     601 AGTCATCCAGGGCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGG 660
Db     601 AGTCATCCAGGGCTCCCAAAGCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGG 660

Qy     661 CAAGTTGGTATCAAGGTGGCCACTGGTCCCCGCGGATGCTGTCCCCCAGCCCAGGAA 720
Db     661 CCAAGTTGGTATCAAGGTGGCCACTGGTCCCCGCGGATGCTGTCCCCCAGCCCAGGAA 720

Qy     721 AGTGCCCTGGCCAGCTGTGGCTCCTCGACCCAAGGCCCGGCTACAGCTTGGGCCCTCTA 780
Db     721 AGTGCCCTGGCCAGCTGTGGCTCCTCGACCCAAGGCCCGGCTACAGCTTGGGCCCTCTA 780

Qy     781 CTAGAAGAAAGGCAGGGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTG 840
Db     781 GCTTCATCAAGGAAAAGCAGGGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTG 840

Qy     841 CCCACACACCCCACCTCCACCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGACCCC 900
Db     841 CCCACACACCCCACCTCCACCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGACCCC 900

Qy     901 TTTCAGCAGAGCCGTGGTGGTTGACACAGCCCGGGACGACTCCTCCTCCACCCAGC 960
Db     901 TTTCAGCAGAGCCGTGGTGGTTGACACAGCCCGGGACGACTCCTCCTCCACCCAGC 960

Qy     961 TACTCGCGCCCTCTGGCAGCGTGGCTTCTCTACACGGCACTCTCCTTGAACTTGTTCC 1020
Db     961 TACTCGCGCCCTCTGGCAGCGTGGCTTCTCTACACGGCACTCTCCTTGAACTTGTTCC 1020

Qy    1021 TACGCAAGGAGGTGTTCTACCCACGGGAGAACTTCAGCATACTACACATCGACCC 1080
Db    1021 TACGCAAGGAGGTGTTCTACCCACGGGAGAACTTCAGCATACTACACATCGACCC 1080

Qy    1081 TCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTATCTGGATTTCCCAGAATG 1140
Db    1081 TCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTATCTGGATTTCCCAGAATG 1140

Qy    1141 AGCGGGGGAAAATGAAAGACCTGCTGGGAGGCTTGGAGGTGGAATTCACCAACTTCCCA 1200
Db    1141 AGCGGGGGAAAATGAAAGACCTGCTGGGAGGCTTGGAGGTGGAATTCACCAACTTCCCA 1200

Qy    1201 CCACCGAAGACAGGGTCAAGAAGGGCATCGTGGTGGCCCACTGGGACAACTGGGCCAATT 1260
Db    1201 CCACCGAAGACAGGGTCAAGAAGGGCATCGTGGTGGCCCACTGGGACAACTGGGCCAATT 1260

Qy    1261 ACTTCTCCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGCGACCTGTTA 1320
Db    1261 ACTTCTCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGCGACCTGTTA 1320

Qy    1321 TGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCAAAGGCTCCGGCCCGACT 1380
Db    1321 TGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCAAAGGCTCCGGCCCGACT 1380

Qy    1381 AGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGTGGAGTGCCGGGCG 1440
Db    1381 AGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGTGGAGTGCCGGGCG 1440

Qy    1441 GCTCCACCCTGGAGCTGTCACTGAAGAGGGAGCAGTGCTGCTCCACACAGCCGGGAA 1500
Db    1441 GCTCCACCCTGGAGCTGTCACTGAAGAGGGAGCAGTGCTGCTCCACACAGCCGGGAA 1500

Qy    1501 GGGCCATGGAGGGGCTGGTTGAGCTATTCCTGAATGACCTAACAAACAGCACTCCGCCTATG 1560
Db    1501 GGGCCATGGAGGGGCTGGTTGAGCTATTCCTGAATGACCTAACAAACAGCACTCCGCCTATG 1560

Qy    1561 TCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCACCGTGGG 1620
Db    1561 TCATCGCCCTGCGCAGCTACATCACTGACAACTGCAGCCTCCTCACCGTGGG 1620

Qy    1621 ACCTCATCAAGCTGCTGCCGGTGGCCACCCTGGAGCCAGGCTGGCAGTTTGCCTGCG 1680
Db    1621 ACCTCATCAAGCTGCTGCCGGTGGCCACCCTGGAGCCAGGCTGGCAGTTTGCCTGCG 1680

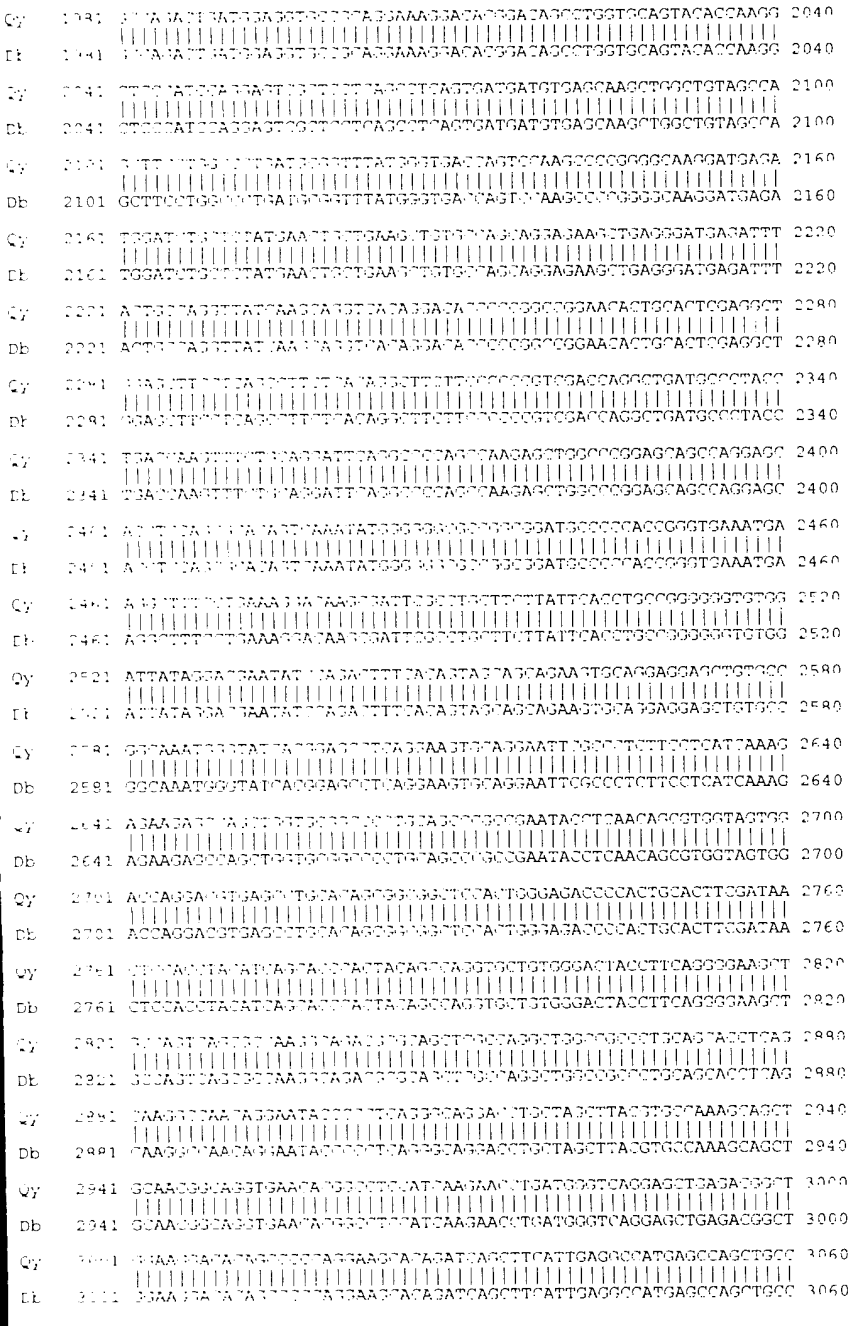
Qy    1681 GGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTGCAGCCCCTGCCGACTTTT 1740
Db    1681 GGGGCCGTTCCGGACTCTTTCCTGCCGACATAGTGCAGCCCCTGCCGACTTTT 1740

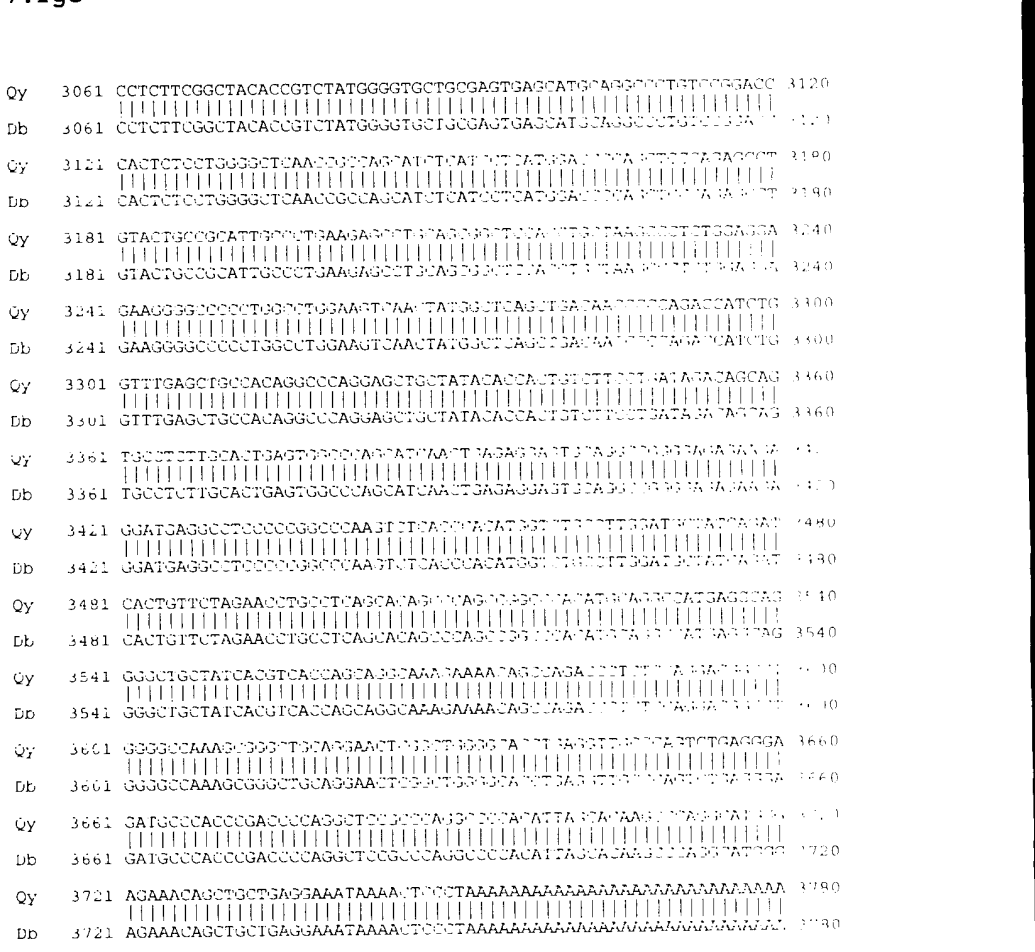
Qy    1741 CCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTCAGCTGTCCAAGGGGAACCTAG 1800
Db    1741 CCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTCAGCTGTCCAAGGGGAACCTAG 1800

Qy    1801 GGCTGGCTCGGTGGGACAGGGCCTCAGAGCGCCCTGCCACCTTGGAGCCAGGCACACA 1860
Db    1801 GGCTGGCTCGGTGGGACAGGGCCTCAGAGCGCCCTGCCACCTTGGAGCCAGGCACACA 1860

Qy    1861 GTGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATCCCTCGGACTCCC 1920
Db    1861 GTGACGACTCGGAGGCCACCAGCCTGTCCTCTGTGGCCTATCCCTCGGACTCCC 1920

Qy    1921 ACAGCTACACCATGCAGGAATTCGCCCGGCGTTACTTCCGAGCCTCCTCTGG 1980
Db    1921 ACAGCTACACCATGCAGGAATTCGCCCGGCGTTACTTCCGAGCCTCCTCTGG 1980
```

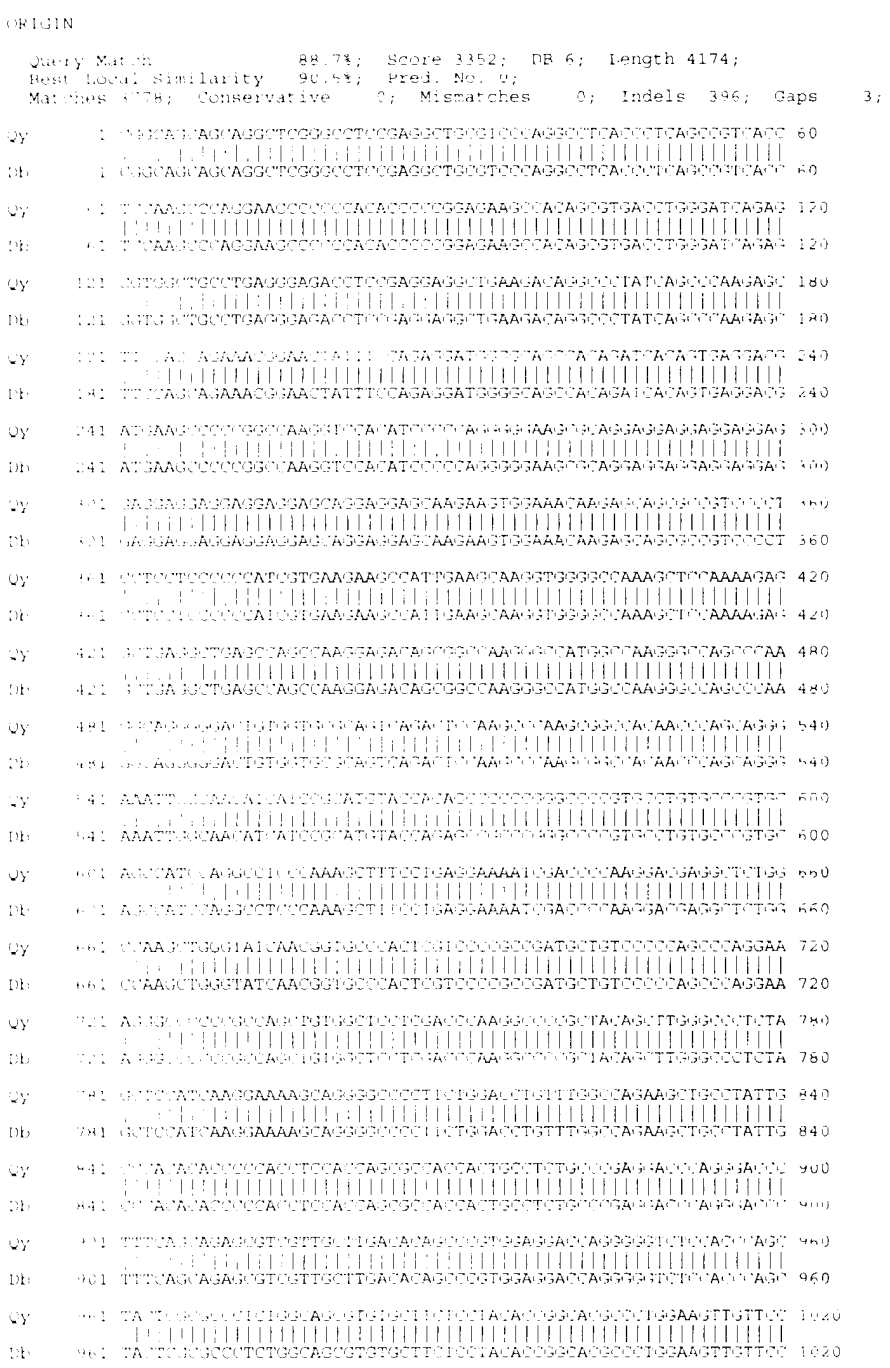


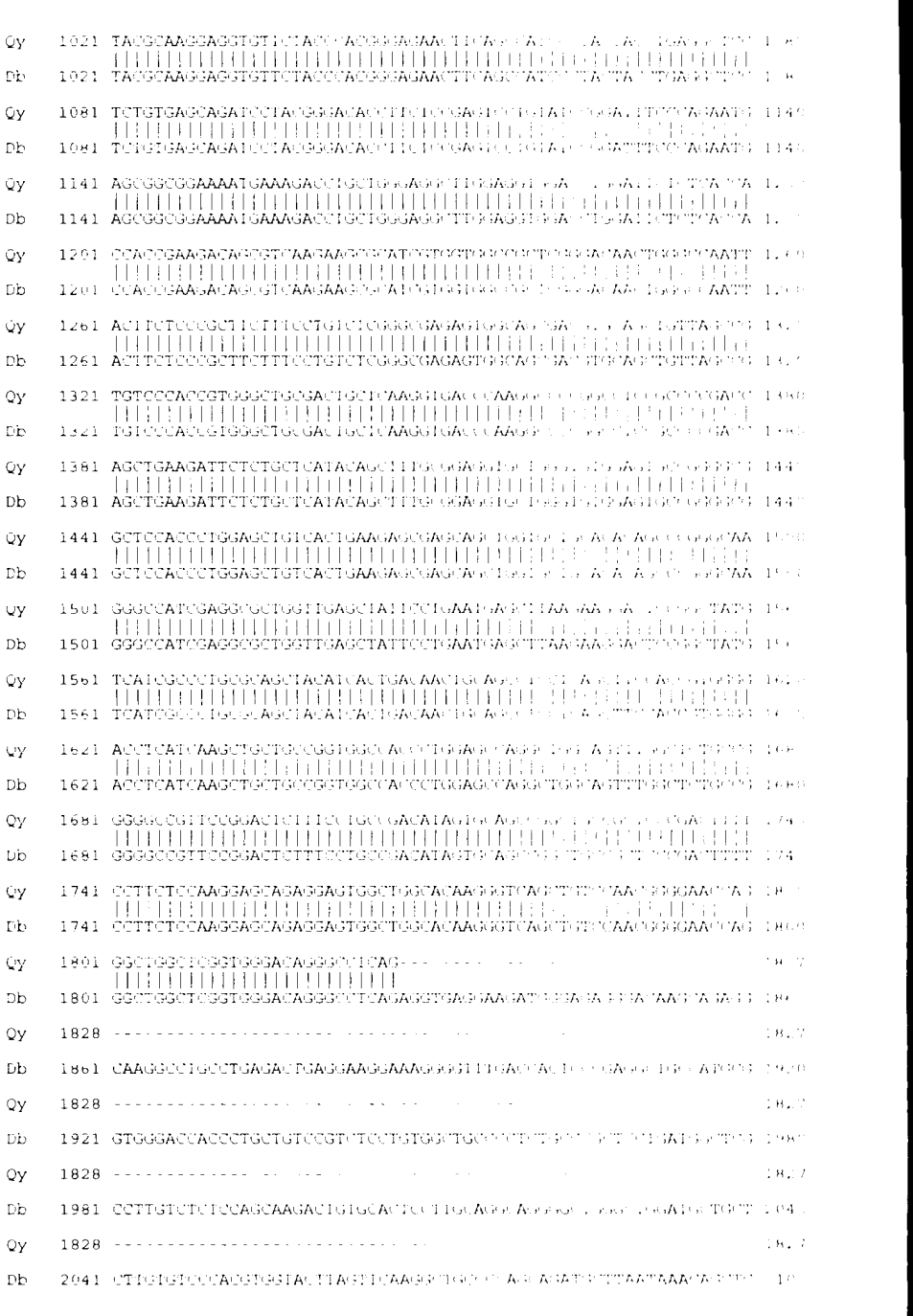


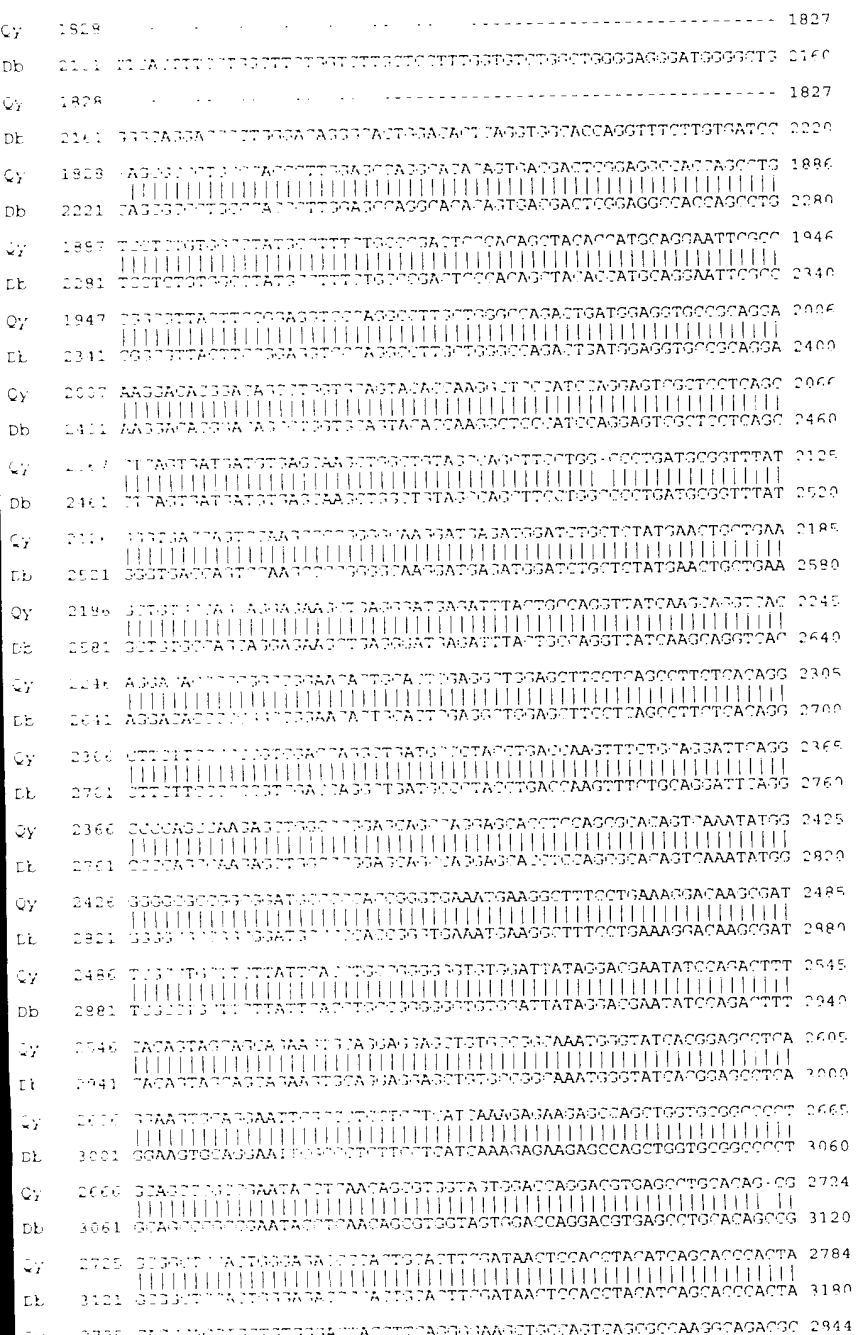
```
RESULT 2
AX253082
LOCUS       AX253082                4174 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 5 from Patent WO0168866
ACCESSION   AX253082
VERSION     AX253082.1  GI:15986279
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4174)
  AUTHORS   Brooks,A.P., Deng,G.G. and Rubanyi,G.M.
  TITLE     Estrogen-regulated unconventional myosin related protein:
            compositions and methods of use
  JOURNAL   Patent: WO 0168866-A 5 20-SEP-2001;
            SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..4174
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Partial DNA sequence for human myosin related
                     protein variant 1 (hMRP1)"
BASE COUNT      901 a   1339 c   1237 g    697 t
```

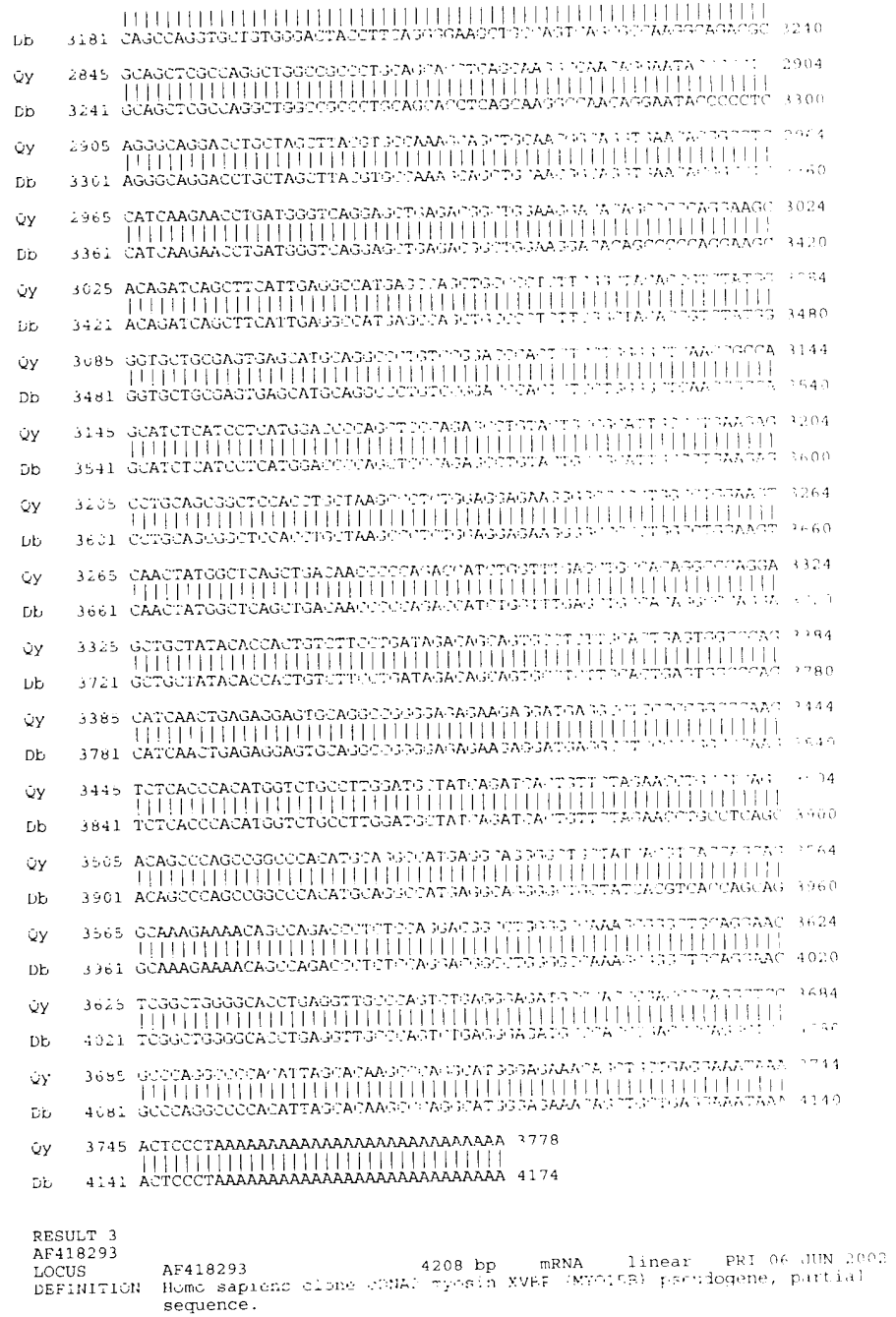
ORIGIN

Query Match 88.7%; Score 3352; DB 6; Length 4174;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 3778; Conservative 0; Mismatches 0; Indels 396; Gaps 3;









RESULT 3
AF418293
LOCUS AF418293 4208 bp mRNA linear PRI 06-JUN-2002
DEFINITION Homo sapiens clone xxNAL myosin XVB (MYO15B) pseudogene, partial sequence.

```
ACCESSION   AF418293
VERSION     AF418293.1  GI:16903816
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4208)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-483 (2001)
  MEDLINE   21968014
   PUBMED   11964073
REFERENCE   2  (bases 1 to 4208)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .4208
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map "17q24-q25"
                     /clone="cDNA2"
                     /tissue_type="kidney; stomach"
     gene            <1. .>4208
                     /gene "MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT      867 a   1386 c   1275 g    680 t
ORIGIN
```

Query Match 81.8%; Score 3167; DB 9; Length 4208;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 3123; Conservative 9; Mismatches 5; Indels 3; Gaps 3;

```
Qy       1 CCCTCAGCAGCAGGCTGGGGCCTCCGAGGCTGGGTCCCAGGCCTCAGCCTCAGCCGTCACC  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     999 CCCCAGCAGCAGGCTGGGGCCTCCGAGGCTGGGTCCCAGGCCTCAGCCTCAGCCGTCACC 1058

Qy      61 TCTAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCCTGACCTGGGATCAGAG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1059 TCTAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCCTGACCTGGGATCAGAG 1118

Qy     121 CCTGGCCTGCTGAGGCAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1119 CCTGGCCTGCTGAGGCAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 1178

Qy     181 TTTCAGCAGAAACGGAACTATTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1179 TTTCAGCAGAAACGGAACTATTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 1238

Qy     241 ATCAAGCCCCCGGCCAAGGTCACATCCCCAGGGGGAAGCCAGGAGGAGGAGGAGGAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1239 ATCAAGCCCCCGGCCAAGGTCACATCCCCAGGGGGAAGCCAGGAGGAGGAGGAGGAG 1298

Qy     301 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGGGCCGTCCCCT 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1299 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGGGCCGTCCCCT 1358

Qy     361 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1359 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 1418

Qy     421 GCTGAGGCTGAGCCAGCCAAGGAGACAGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1419 GCTGAGGCTGAGCCAGCCAAGGAGACAGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 1478

Qy     481 GCTAGGGGAGTGTGGTGGCAGTCAGACTCCAAGCCCAAGCGGCCACAAGCCAGCAGG 539
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1479 GCTAGGGGAGTGTGGTGGCAGTCAGACTCCAAGCCCAAGCGGCCACAAGCCAGCAGG 1538

Qy     540 GAAATTGGCAACATCATCGGCATGTATCAGAGCCGCCGCCC                    579
           |||||||||||||||||||||||||||||||||||||||||||                    
Db    1539 GAAATTGGCAACATCATCGGCATGTATCAGAGCCGCCGCCCTTTCTTTCTTTTTTTTTTT 1598

Qy     580 CAGCCATCCAGGGCTCCCAAAGTTTCCTGAGAAAATCAAGT                    619
           |||||||||||||||||||||||||||||||||||||||||||                    
Db    1599 CAGCCATCCAGGGCTCCCAAAGTTTCCTGAGAAAATCAAGTTTTTTTTTTTTTTTTTTT 1658

Qy     620 GCCAAGTTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTTTTCCCCAGCCCAGGA 679
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1659 GCCAAGTTGGGTATCAACGGTGCCCACTCGTCCCCGCCGATGCTTTTCCCCAGCCCAGGA 1718

Qy     680 AAGGGCCCCCCGCCAGCTGTGGCTCCTCGACCCAAGGGCCCCCACAGTTGGGCCCTCT 739
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1719 AAGGGCCCCCCGCCAGCTGTGGCTCCTCGACCCAAGGGCCCCCACAGTTGGGCCCTCT 1778

Qy     740 AGCTCCACCAAGGAAAAGCAGGGGCCCCTCCTCGGACCCGCTCCCACAAAGCCCCTATT 799
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1779 AGCTCCACCAAGGAAAAGCAGGGGCCCCTCCTCGGACCCGCTCCCACAAAGCCCCTATT 1838

Qy     800 GCCCACACACCCCCACCTCCACCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCT 859
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1839 GCCCACACACCCCCACCTCCACCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCT 1898

Qy     860 CTTTCAGCAGAGGGTCGTGGTTGACACAGCCCGTGGAGGACCAGGGGTCTCCACCCAG 919
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1899 CTTTCAGCAGAGGGTCGTGGTTGACACAGCCCGTGGAGGACCAGGGGTCTCCACCCAG 1958

Qy     920 CTACTCGGCGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGTACTCCTTGGAAGTTTTT 979
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1959 CTACTCGGCGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGTACTCCTTGGAAGTTTTT 2018

Qy     980 CTACGGCAAGGAGGTGTTCTACCTACCGGAGAACTCCACGCCAGCCCTCACCTGGATCCTC 1039
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2019 CTACGGCAAGGAGGTGTTCTACCTACCGGAGAACTCCACGCCAGCCCTCACCTGGATCCTC 2078

Qy    1040 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTTTATCCCAGAATTCCTCAAT 1099
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2079 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTTTATCCCAGAATTCCTCAAT 2138

Qy    1100 GAGCGGCGGGAAAATGAAAGACCTGCTGGGAGGCTTGGAGGTGGACCTGGATTCTCTCACC 1159
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2139 GAGCGGCGGGAAAATGAAAGACCTGCTGGGAGGCTTGGAGGTGGACCTGGATTCTCTCACC 2198

Qy    1160 ACCACCGAAGACAGCGTCAAGAAGCGCATCGTGTGGCGCCTCCGGACAACTGGGCCAAT 1219
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2199 ACCACCGAAGACAGCGTCAAGAAGCGCATCGTGTGGCGCCTCCGGACAACTGGGCCAAT 2258

Qy    1220 TACTTCTCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGGAACTGTTA                1269
           |||||||||||||||||||||||||||||||||||||||||||||||||                
Db    2259 TACTTCTCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGGAACTGTTATTTTTTTTTT 2318

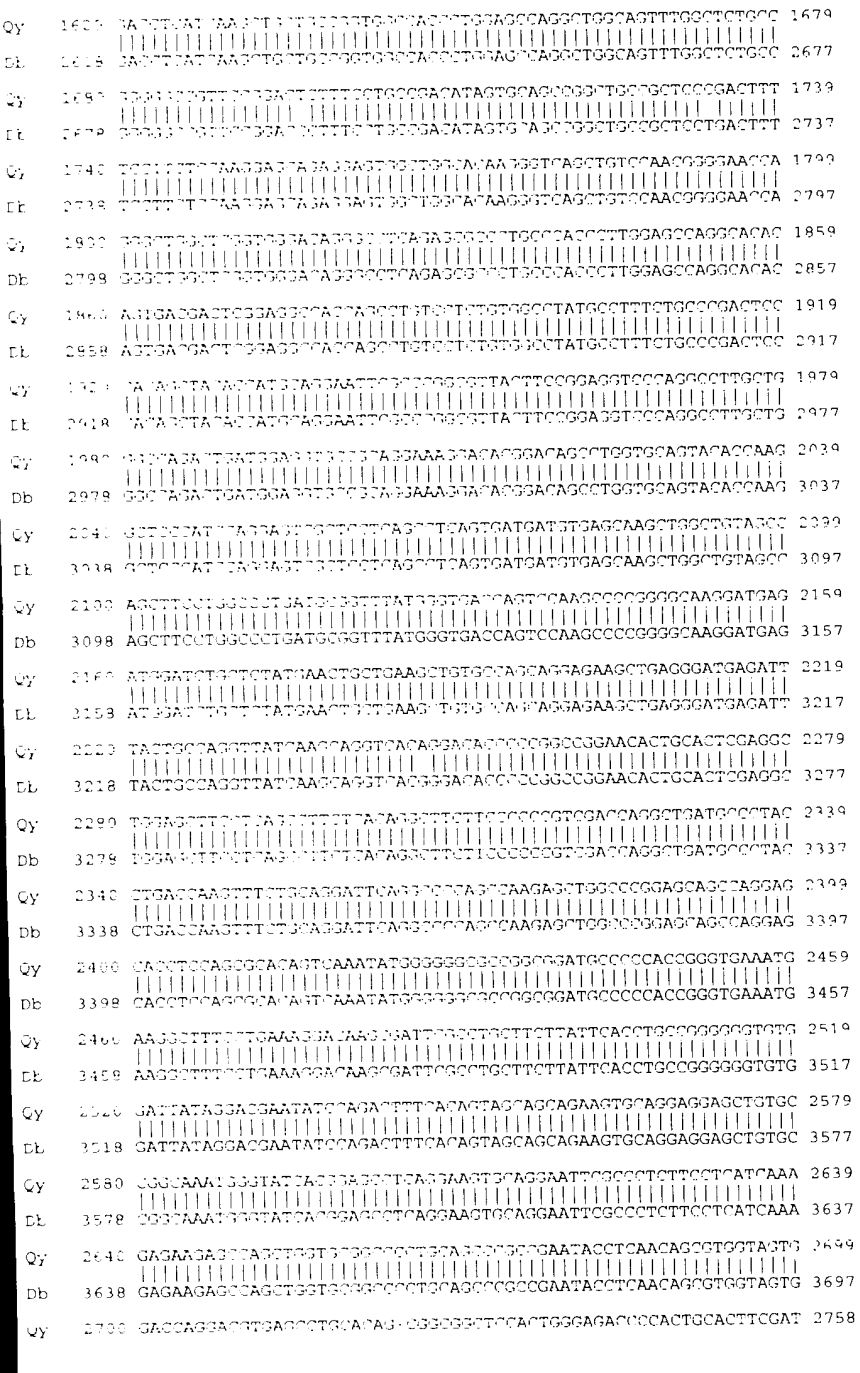
Qy    1270 GTGTCCCACCGTGGGCTGTGACTGCTCAAGGTGACCCAAGG                       1309
           |||||||||||||||||||||||||||||||||||||||||||                       
Db    2319 GTGTCCCACCGTGGGCTGTGACTGCTCAAGGTGACCCAAGGTTTTTTTTTTTTTTTTTT 2378

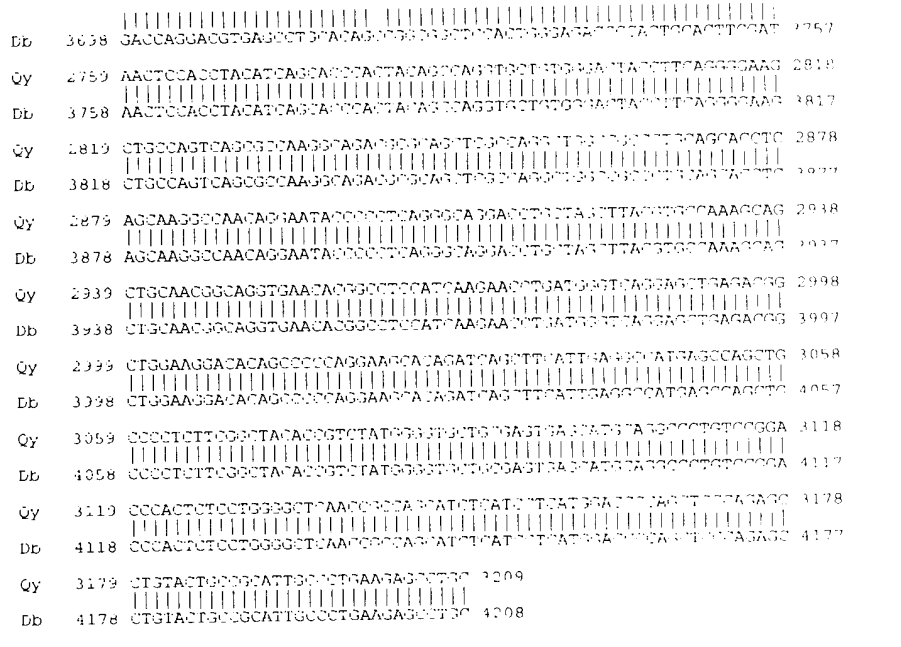
Qy    1310 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGGCGAGGCCCGGGC 1369
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2379 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGGCGAGGCCCGGGC 2438

Qy    1370 GGCTCCACCCTGGAGCTGTCACTGAAGAGCAGCAGCAGTGGGTGCACACAGCCCGGGCA 1429
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2439 GGCTCCACCCTGGAGCTGTCACTGAAGAGCAGCAGTGGGTGCACACAGCCCGGGCA 2498

Qy    1430 AGGGCCATCGAGGGCCTGGTTGAGCTATTCCTGAATGACTTTAAGAAAGTCCCCTTTTAT 1489
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2499 AGGGCCATCGAGGGCCTGGTTGAGCTATTCCTGAATGACTTTAAGAAAGTCCCCTTTTAT 2558

Qy    1490 GTCATCGGCCTGCGCAGCTACATCACTGACAACTGCAGCCCTTTTAGTTTTAAGTTTGTG 1549
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2559 GTCATCGGCCTGCGCAGCTACATCACTGACAACTGCAGCCCTTTTAGTTTTAAGTTTGTG 2618
```

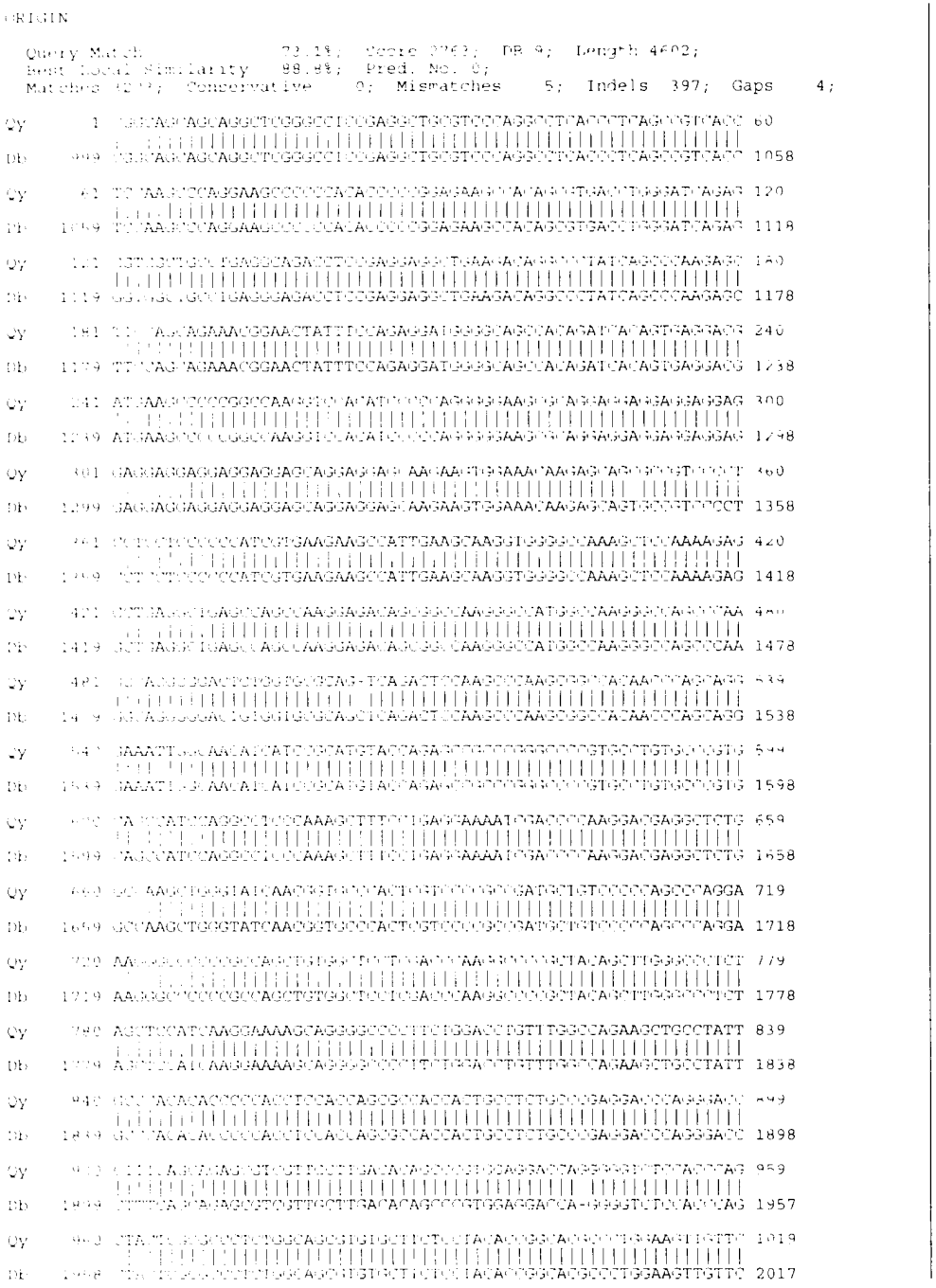


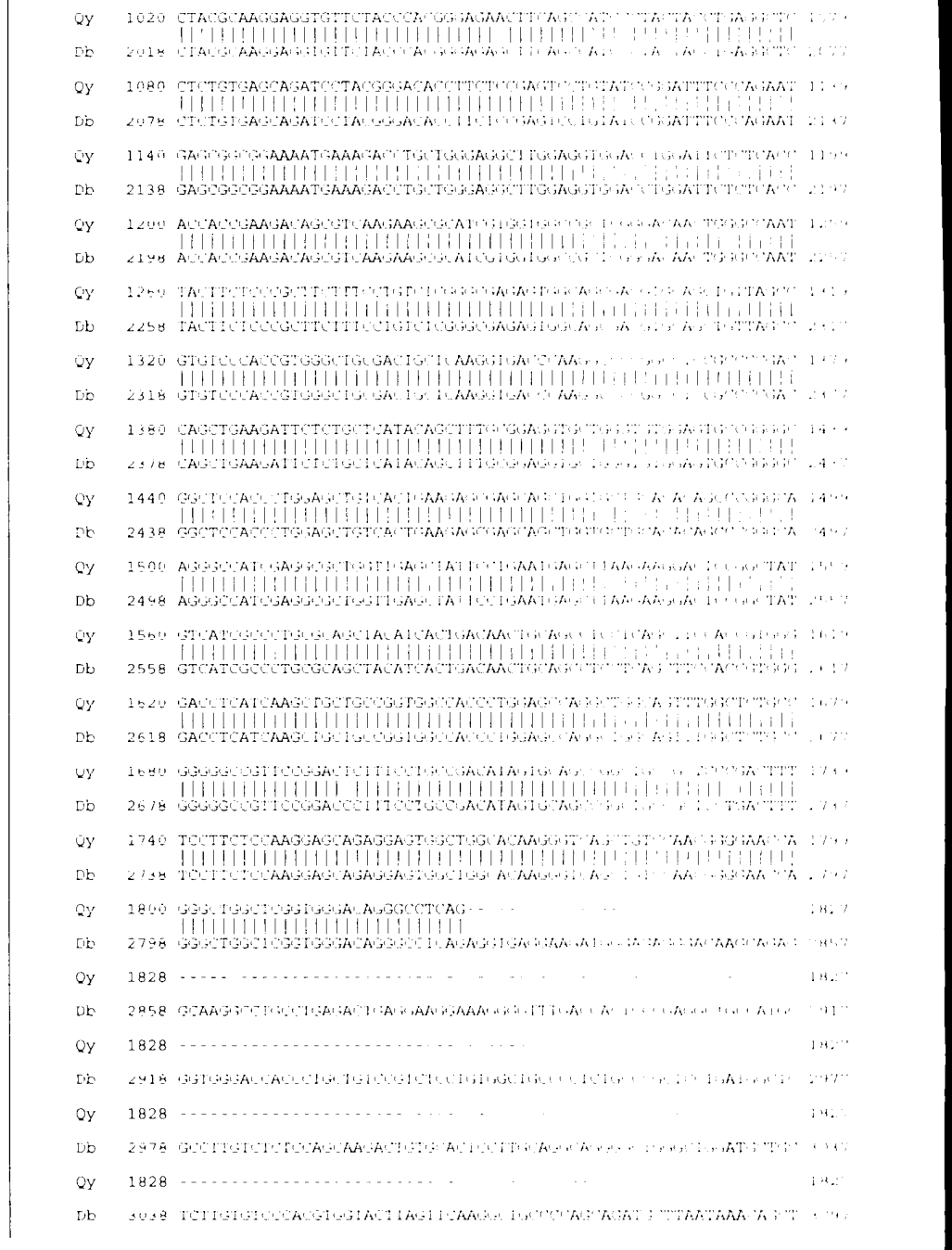


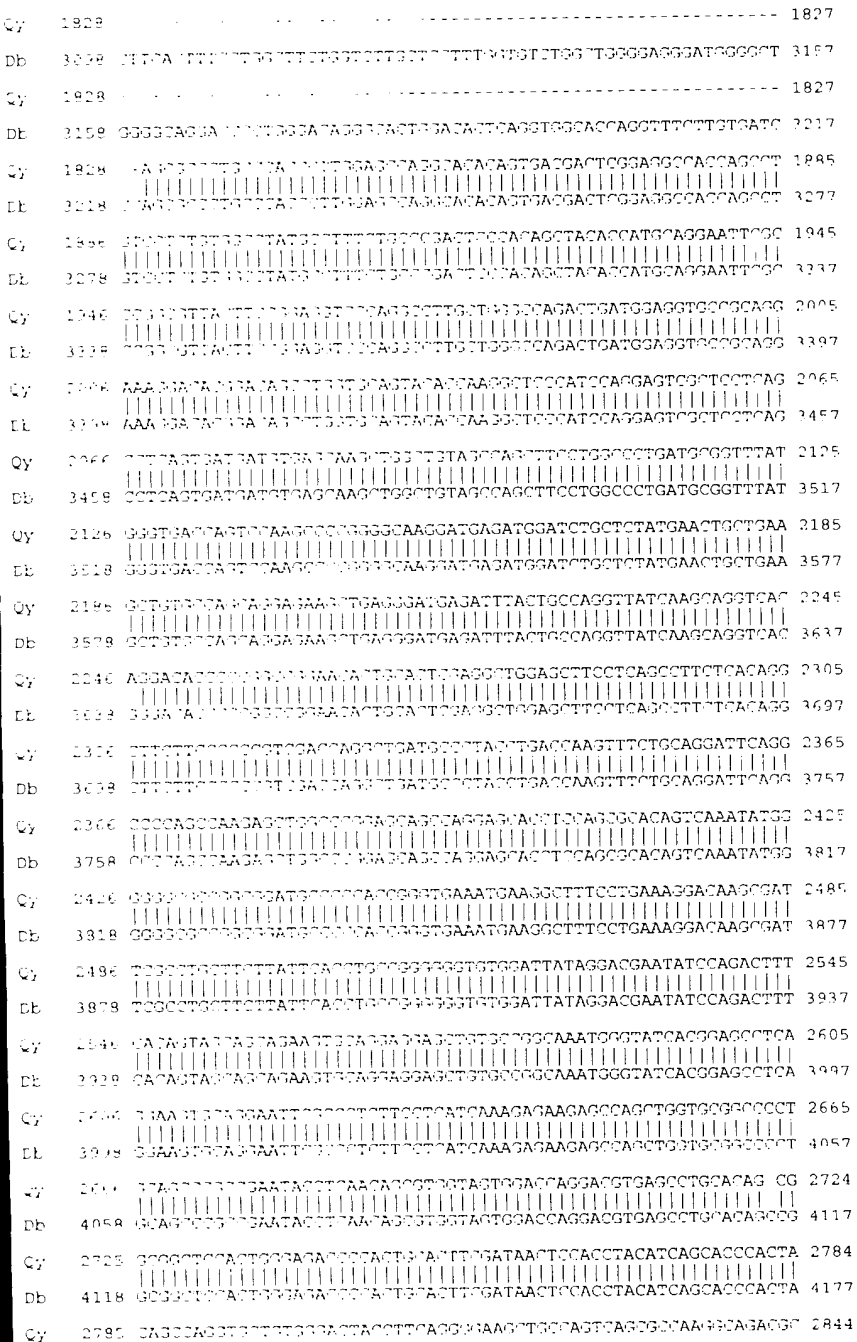
```
RESULT 4
AF418292
LOCUS       AF418292                4602 bp    mRNA    linear   PRI 06-JUN-2002
DEFINITION  Homo sapiens clone cDNA1 myosin XVBP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418292
VERSION     AF418292.1  GI:16303815
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4602)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVBP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477-484 (2001)
  MEDLINE   21960014
   PUBMED   11964073
REFERENCE   2  (bases 1 to 4602)
  AUTHORS   Boger,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2001) Laboratory of Molecular Genetics, National
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..4602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24-q25"
                     /clone="cDNA1"
                     /tissue_type="kidney; stomach"
     gene            <1..>4602
                     /gene="MYO15B"
                     /note="myosin XVBP"
                     /pseudo
BASE COUNT      927 a   1495 c   1405 g    775 t
```

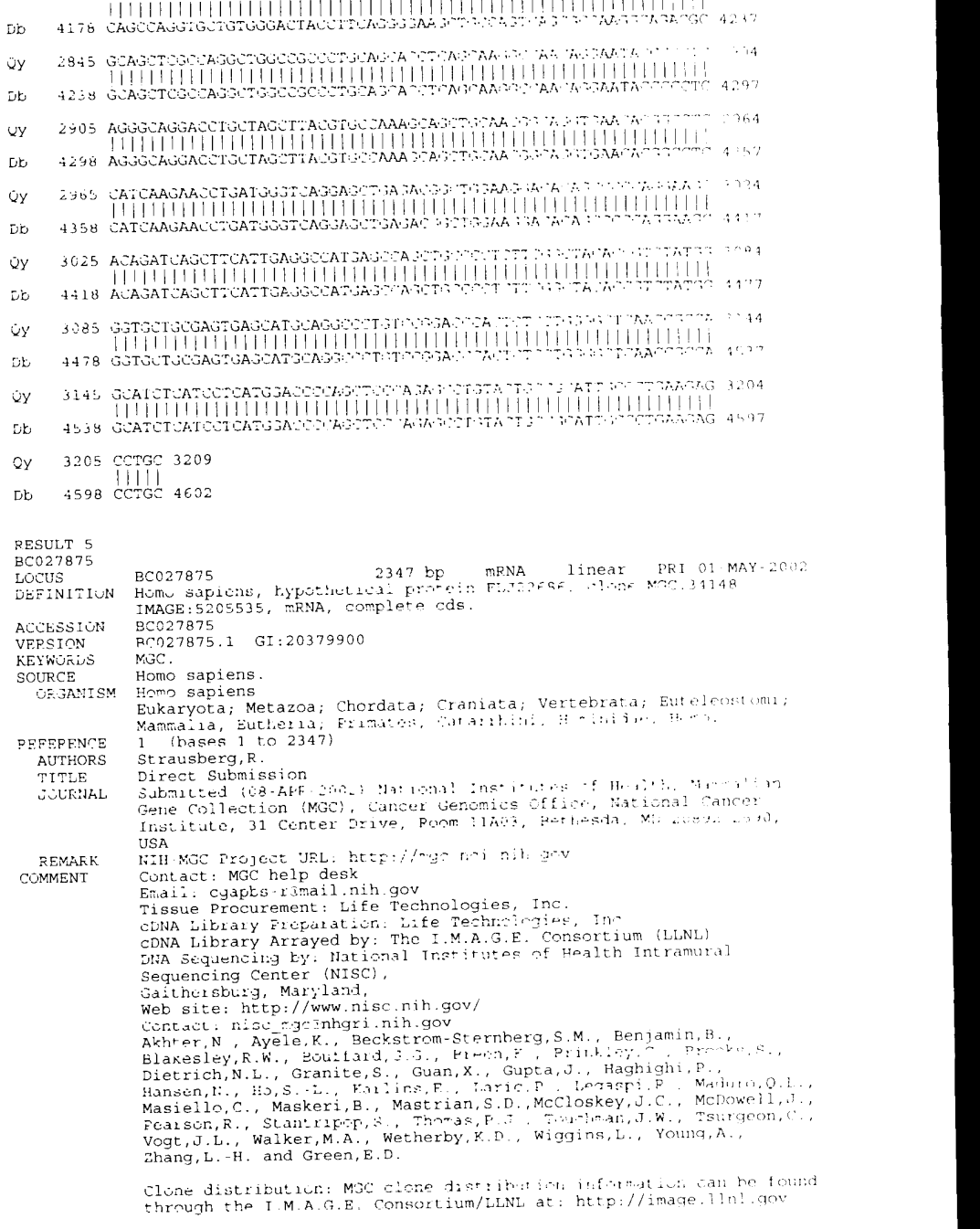
ORIGIN

Query Match 72.1%; Score 2763; DB 9; Length 4602;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 4213; Conservative 0; Mismatches 5; Indels 397; Gaps 4;









RESULT 5
BC027875
LOCUS BC027875 2347 bp mRNA linear PRI 01-MAY-2002
DEFINITION Homo sapiens, Hypothetical protein FLJ22686, clone MGC:34148 IMAGE:5205535, mRNA, complete cds.
ACCESSION BC027875
VERSION BC027875.1 GI:20379900
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2347)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r3mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland,
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brooks,S., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

```
            Series: IRAK Plate: 49 Row: g Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 13376444.
FEATURES             Location/Qualifiers
     source          1. .2347
                     /organism="Homo sapiens"
                     /db_xref="LocusID:80022"
                     /db_xref "taxon:9606"
                     /clone="MGC:34148 IMAGE:5205535"
                     /tissue_type="Lung, Spleen, fetal, pooled"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note "Vector: pCMV CFGST6"
     CDS             418. .1881
                     /codon_start=1
                     /product "hypothetical protein FLJ27686"
                     /protein_id="AAH27875.1"
                     /db_xref "GI:22779961"
                     /translation="MQEFARPYFPPSQALLCQTDWAAVKPIDSLVQYTKAPIQESLL
                     SLSDVSPKLAVASFIACMPFMCPQSFPRCKDEMDLLYELIFLTQKEKISEIYTQVIK
                     QVTGHPRPFHCTPGWSFLSLLTGFFPPSIPLMPYLTKFLQDSGPSQELARSSQEHLQR
                     TVFYCIFRFMYPPTPEWFAIIPSLAIRLLTHLIPCVLYFIDLGFIVAAFVLFELFS
                     MGITFPQEVQEFALFLIKEKSQLVPPLQPAEYLSSVVVDQDVSILHSPPLHWETPLHFD
                     NCTYICTHTGQVLWDTLLGFLIGLAFALQLARLAALALPLFANENTFSSLDLLAYYI
                     KQLQRQVNTASIPNLMSQELRRLEGHSPQEAQISFTEAMSQLPLFGYTVYGVLPVSMQ
                     ALSGPTLLGLNPQHLILMCTSSLCLYCPIALPSLQPLHLLSPLEEPPPHSLEVNYGSA
                     DNPQTIWFELPQAQELLYTTVFLIDSSASCTEWPSIN"
BASE COUNT      574 a      619 c      654 g      400 t
ORIGIN

  Query Match          51.8%;  Score 1937;  DB 9;  Length 2347;
  Best Local Similarity   98.7%;  Pred. No. 0;
  Matches 1693;  Conservative    0;  Mismatches   25;  Indels    1;  Gaps    1;

Qy    1793 GGAACAAGGGCTGGCTCGGTGGGACAGGGCCTCAGAGCGCCCTGCCCACCCTTGGAGCCA 1852
Db     279 GGACACTCAGGTGGCACCAGGTTTCTTGTGAACCCCTGCCCACCCTTGGAGCCA 338

Qy    1853 GCACACAGTGACGACTCGGAGGCCACAGCGTGTGCTCTGTGGCCTATGCCTTTCTGCC 1912
Db     339 GCACACAGTGACGACTCGGAGGCCACAGCGTGTGCTCTGTGGCCTATGCCTTTCTGCC 398

Qy    1913 CGACTCCCACAGCTACACCATGCAGGAATTCGCCCCGGCGTTACTTCCGGAGGTCCCAGGC 1972
Db     399 CGACTCCCACAGCTACACCATGCAGGAATTCGCCCCGGCGTTACTTCCGGAGGTCCCAGGC 458

Qy    1973 TTTGCTGGGCCAGACTGATGGAGGTGCCGCAGGAAAGGACACGGACAGCCTGGTGCAGTA 2032
Db     459 CTTGCTGGGCCAGACTGATGGAGGTGCCGCAGGAAAGGACACGGACAGCCTGGTGCAGTA 518

Qy    2033 TACAAGGCTCCCATCCAGGAGTCGCTCCTCAGCCTCAGTGATGATGTGAGCAAGCTGGC 2092
Db     519 CACCAAGGCTCCCATCCAGGAGTCGCTCCTCAGCCTCAGTGATGATGTGAGCAAGCTGGC 578

Qy    2093 TGTAGCCAGCTTCCTGGCCCTGATGGGGTTTATGGGTGACCAGTCCAAGCCCCGGGGCAA 2152
Db     579 TGTAGCCAGCTTCCTGGCCCTGATGGGGTTTATGGGTGACCAGTCCAAGCCCCGGGGCAA 638

Qy    2153 GGATCAGATGGATCTCCTCTATGAACTGCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGA 2212
Db     639 GGATCAGATGGATCTCCTCTATGAACTGCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGA 698

Qy    2213 TGACATTACTGCCAGGTTATCAAGCAGGTCACAGGACACCCCCGGGCCCGAACACTGTAC 2272
Db     699 TGACATTACTGCCAGGTTATCAAGCAGGTCACAGGACACCCCCGGGCCCGAACACTGTAC 758

Qy    2273 TTCAAGGGGAAGCTTCCTCAGCCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGAT 2332
Db     759 TTGAGGCTGGAGCTTCCTCAGCCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGAT 818

Qy    2333 GCCCTACCTGACCAAGTTTCTGCAGGATTCAGGCCCCAGCCAAGAGCTGGCCCGGAGCAG 2392
Db     819 GCCCTACCTGACCAAGTTTCTGCAGGATTCAGGCCCCAGCCAAGAGCTGGCCCGGAGCAG 878

Qy    2393 CCAGAAGTACCTCCAGCGCACAGTCAAATATGGGGGCCCCCTTGACTTCCCCATCCTT 2452
Db     879 CCAGAAGTACCTCCAGCGCACAGTCAAATATGGGGGCCCCCTTGACTTCCCCATCCTT 938

Qy    2453 TGAAATGAAGGCTTTCCTCAAAGCACAAGCGATCCCAGTGCTTATTTACCTGCCCC 2511
Db     939 TGAAATGAAGGCTTTCCTCAAAGCACAAGCGATCCCAGTGCTTATTTACCTGCCCC 998

Qy    2513 GGGTGTGGATTATAGGACGAATATCCAGACTTTCACAGTAGCAGCAAAGTGCAGGAGGA 2572
Db     999 GGGTGTGGATTATAGGACGAATATCCAGACTTTCACAGTAGCAGCAAAGTGCAGGAGGA 1058

Qy    2573 GCTGTGCCGGCAAATGGGTATCACGGAGCCTCAGGAAGTCCAAATTGCCCCTTTCCT 2632
Db    1059 GCTGTGCCGGCAAATGGGTATCACGGAGCCTCAGGAAGTCCAAATTGCCCCTTTCCT 1118

Qy    2633 CAGCAAAGAGAAGAGGCTAGCCTGCCGCGGCGCCCTTGCAGCCCCCCAAATACTCTAGCAGCT 2692
Db    1119 CAGCAAAGAGAAGAGGCTAGCCTGCCGCGGCGCCCTTGCAGCCCCCCAAATACTCTAGCAGCT 1178

Qy    2693 GGTACGAGCTACAGGACTCTGAGTCTGCAGCCCCCAGCCTCTACAGCAGCTACCCTGCA 2752
Db    1179 GGTACGAGCTACAGGACTCTGAGTCTGCAGCCCCCAGCCTCTACAGCAGCTACCCTGCA 1238

Qy    2753 CTTCGATAACTCTGCAGCTACATCAGCCAGCCCACTACAGCAGCAGCCCTCCCCAGTGCA 2812
Db    1239 CTTCGATAACTCTGCAGCTACATCAGCCAGCCCACTACAGCAGCAGCCCTCCCCAGTGCA 1298

Qy    2813 GGGGAAGCTGCCAGTCAGCGCCAAGGCAGACGCGCAGCTCGCCAGGCTGCCCTGCA 2872
Db    1299 GGGGAAGCTGCCAGTCAGCGCCAAGGCAGACGCGCAGCTCGCCAGGCTGCCCTGCA 1358

Qy    2873 GCACCTCAGCAAGGCCAACAGGAATACCCCCTCAGCGAGGAGCTCGCTAGCTTACGTGCC 2932
Db    1359 GCACCTCAGCAAGGCCAACAGGAATACCCCCTCAGCGAGGAGCTCGCTAGCTTACGTGCC 1418

Qy    2933 AAAGCAGCTGCAAGGGCAGGTGAACACGGCCTCCATCAAGAACTTCGATGTGTCAGAGCT 2992
Db    1419 AAAGCAGCTGCAAGGGCAGGTGAACACGGCCTCCATCAAGAACTTCGATGTGTCAGAGCT 1478

Qy    2993 GAGACGGCTGGAAGGACACAGCCCCCAGGAAGCACAGATCAGTTTATTGAGGCCATGAG 3052
Db    1479 GAGACGGCTGGAAGGACACAGCCCCCAGGAAGCACAGATCAGTTTATTGAGGCCATGAG 1538

Qy    3053 CCAGCTGCCCCTCTTCGGCTACACCGTCTATGGGGTGCTGCTGCAGTGACATCATGAGCCT 3112
Db    1539 CCAGCTGCCCCTCTTCGGCTACACCGTCTATGGGGTGCTGCTGCAGTGACATCATGAGCCT 1598

Qy    3113 GTCCGGACCCACTCTCCTGGGGCTCAACCGCCAGCATCTCATCTCATGGACCCAGCTC 3172
Db    1599 GTCCGGACCCACTCTCCTGGGGCTCAACCGCCAGCATCTCATCTCATGGACCCAGCTC 1658

Qy    3173 CCAGAGCCTGTACTGCCGCATTGCCCTGAAGAGCCTGAAGGCCTAAGCCTAAGCCC 3232
Db    1659 CCAGAGCCTGTACTGCCGCATTGCCCTGAAGAGCCTGAAGGCCTAAGCCTAAGCCC 1718

Qy    3233 TCTGGAGGAGAAGGGGCCCCCTGGCCTGGAAGTCAACTATGCCCAACCTAACCCCA 3292
Db    1719 TCTGGAGGAGAAGGGGCCCCCTGGCCTGGAAGTCAACTATGCCCAACCTAACCCCA 1778

Qy    3293 GACCATCTGGTTTGAGCTGCCACAGGCCCAGGAGCTGCTATACACCACTGTCTTCTTGATAT 3352
Db    1779 GACCATCTGGTTTGAGCTGCCACAGGCCCAGGAGCTGCTATACACCACTGTCTTCTTGATAT 1838

Qy    3353 AGACAGCAGTGCCTCTTGCACTGAGTGGCCCAGCATCAACTGACAGAGAGTGCAGGCCGGG 3412
Db    1839 AGACAGCAGTGCCTCTTGCACTGAGTGGCCCAGCATCAACTGACAGAGAGTGCAGGCCGGG 1898

Qy    3413 GAGAGAAGAGGATGAGGCCTCCCCTGGCCTAAGTCTCAACCACATGTCTCTGCTTGGATG 3472
Db    1899 GAGAGAAGAGGATGAGGCCTCCCCTGGCCTAAGTCTCAACCACATGTCTCTGCTTGGATG 1958
```

```
Qy    3472 CTATCAGATCACTGTTCTAGAACCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCC 3531
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1959 CTATCAGATCACTGTTCTAGAACCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCC 2018

Qy    3532 ATGAGGCAGGGGCTGGTATCACGTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTCCA 3591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2019 ATGAGGCAGGGGCTGGTATCACGTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTCCA 2078

Qy    3592 GGACGGCCTCGGGCCAAAGCGGGCTGCAGGAACTGGCTGGGGCACCTGAGGTTGCCCAG 3651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2079 GGACGGCCTCGGGCCAAAGCGGGCTGCAGGAACTGGCTGGGGCACCTGAGGTTGCCCAG 2138

Qy    3652 TGTGAGGGAGATGCCCACCCTGAGCCCAGGCTCCGCCCAGGCCCCACATTAGCACAAGCCC 3711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2139 TGTGAGGGAGATGCCCACCCTGAGCCCAGGCTCCGCCCAGGCCCCACATTAGCACAAGCCC 2198

Qy    3712 AGGCATGGGAGAAAAAAGTGTGAGGAAATAAAACTCCCTAAAAAAAAAAAAAAAAAAAAA 3771
           ||||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db    2199 AGGCATGGGAGAAAAAGTGCTGTGAGGAAATAAAACTCCCTGGAGAAAAAAAAAAAAAAAA 2258

Qy    3772 AAAAAAAAA 3780
           |||||||||
Db    2259 AAAAAAAAA 2267
```

RESULT 6
AK026339
LOCUS AK026339 1891 bp mRNA linear PRI 09-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22686 fis, clone HSI10987.
ACCESSION AK026339
VERSION AK026339.1 GI:10439173
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens human small intestine cDNA to mRNA, clone_lib:HSI clone:HSI10987.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1891)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

```
FEATURES             Location/Qualifiers
     source          1..1891
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HSI10987"
                     /tissue_type="human small intestine"
                     /clone_lib="HSI"
                     /note="cloning vector pME18SFL3"
     CDS             29..1492
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15453.1"
                     /db_xref="GI:10439174"
                     /translation="MQEPAPPYFPPSQALLGQTDGGAAGKDTDSLVQYTKAPIQESLL
                     SLSDDVSKLAVASFLALMPFMGDQSKPPGPTEMDLLYELLRLCQDEKLRDEINQVIK
                     QVTGHPRPEHCTRGWSFLSLLTGFFPPSTRLMPYLTKFLQDSGPSQELARSGQEHLQR
                     TVKYGGRRRMPPPGEMKAFLPGQAIPLLLIHLPGGVDYRTNIQTFTVAAEVQPELCRQ
                     MGITEPQEVQEFALFLIKEKSQLVRPLQPAECLNSVVVDQDVSLHSRRLHWETPLHFD
                     NSTYISTHYSQVLWDYIQGKLPVSAKADAQLAFLAALQHLSKANFNTPGGQNLAYYP
                     KQLQRQVNTASIKNLMGQELKRLEGHSPQEAQISFIEAMSQLFLPGFTVTGVLRVCMQ
                     ALSGPTLLGLNPQHLILMDPSSQSLYCPIAIKSIQPLHLLSPLEPKGPPGLFVNVGSA
                     DNPQTIWFELPQAQELLYTTVFLIDSSASCTEWPSIN"
```

BASE COUNT 454 a 595 c 536 g 306 t
ORIGIN

```
Query Match             49.1%;  Score 1856.4;  DB 9;  Length 1891;
Best Local Similarity   99.6%;  Pred. No. 0;
Matches 1871;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy    1904 CTTTCTGCCCGACTCCCACAGCTACACCATGCAGGAATTCGGCGGGGGTTACTTCCGGAG 1963
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTTTCTGCCCGACTCCCACAGCTACACCATGCAGGAATTCGGCGGGGGTTACTTCCGGAG 60

Qy    1964 GTCCCAGGCCTTGCTGGGCCAGACTGATGGAGGTGCCGTAGGAAAGGACAGGGACAGCCT 2023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTCCCAGGCCTTGCTGGGCCAGACTGATGGAGGTGCCGTAGGAAAGGACAGGGACAGCCT 120

Qy    2024 GGTGCAGTACACCAAGGCTCCCATCCAGGAGTCGCTCCTCCAGCCTCAGTGATGATGTGAG 2083
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGTGCAGTACACCAAGGCTCCCATCCAGGAGTCGCTCCTCCAGCCTCAGTGATGATGTGAG 180

Qy    2084 CAAGCTGGCTGTAGCCAGCTTCCTGGGCCTGATGCCGTTTATGGGTGACCAGTCCAAGCCC 2143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGCTGGCTGTAGCCAGCTTCCTGGGCCTGATGCCGTTTATGGGTGACCAGTCCAAGCCC 240

Qy    2144 CCGGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAAGCTTGTGCCAGCAGGAGAA 2203
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCGGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAAGCTTGTGCCAGCAGGAGAA 300

Qy    2204 GCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGGTCACAGGATACCCCGACGGGGGA 2263
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGGTCACAGGATACCCCGACGGGGGA 360

Qy    2264 ACACTGCACTCGAGGCTGGAGCTTGCTCAGCCTTCTCACAGGTTTCTTCGGCGGCTGGAC 2323
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACACTGCACTCGAGGCTGGAGCTTGCTCAGCCTTCTCACAGGTTTCTTCGGCGGCTGGAC 420

Qy    2324 CAGGCTGATGCCCTACCTGACCAAGTTTCTGCAGGATTCAGGCCCTAGCCAAGAGCTGGC 2383
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAGGCTGATGCCCTACCTGACCAAGTTTCTGCAGGATTCAGGCCCTAGCCAAGAGCTGGC 480

Qy    2384 CCGGAGCAGCCAGGAGCACCTCCAGCGCACAGTCAAATATGGGGGGCGCGGGGGGATGCC 2443
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCGGAGCAGCCAGGAGCACCTCCAGCGCACAGTCAAATATGGGGGGCGCGGGGGGATGCC 540

Qy    2444 CCCACCGGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGATTCGCCTGCTTCTTATTCA 2503
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCCACCGGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGATTCGCCTGCTTCTTATTCA 600

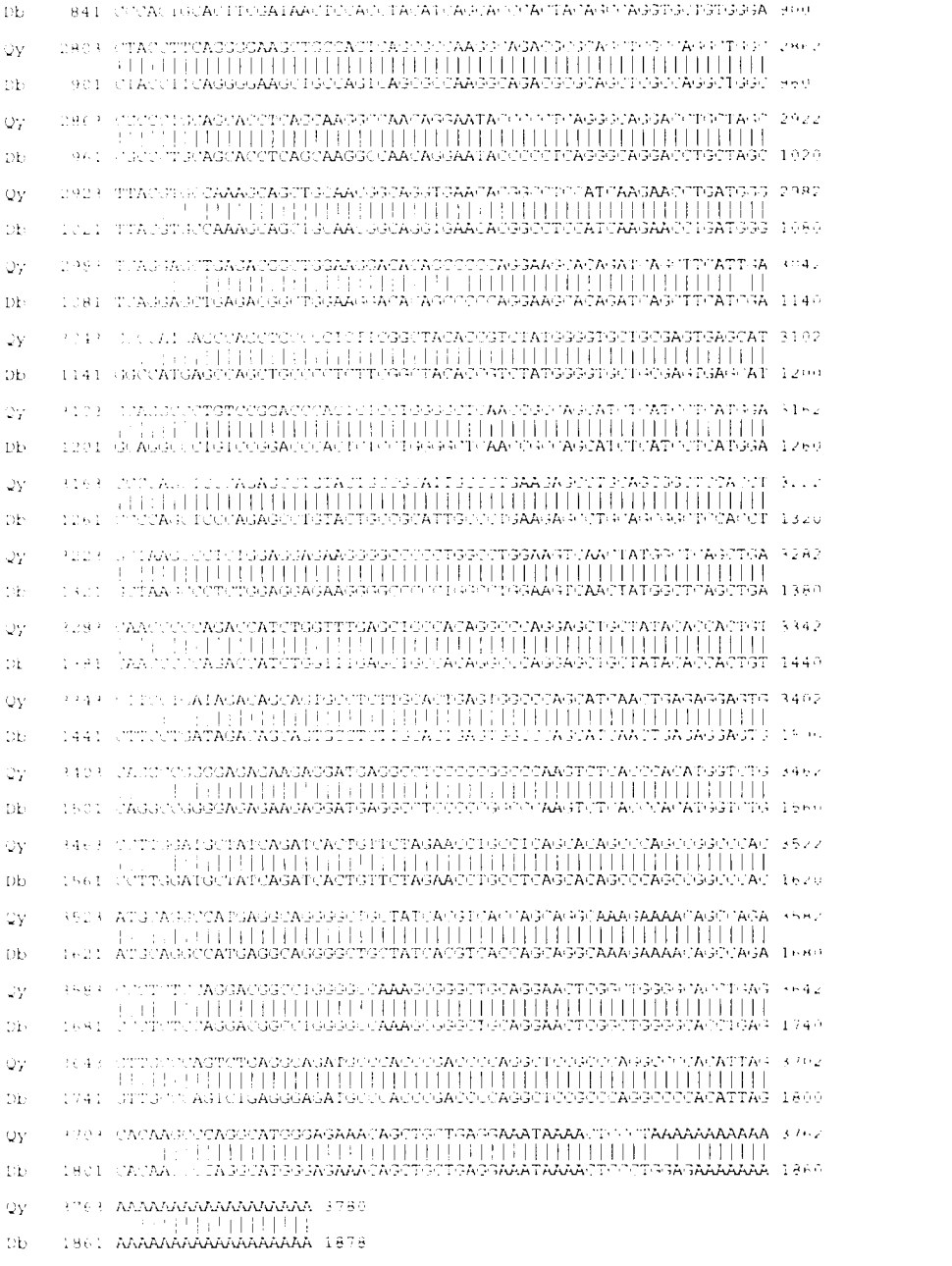
Qy    2504 CCTGCCGGGGGGGTGTGGATTATAGGACGAATATCCAGACTTTCACAGTAGCAGCAGAAGT 2563
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CCTGCCGGGGGGGTGTGGATTATAGGACGAATATCCAGACTTTCACAGTAGCAGCAGAAGT 660

Qy    2564 GCAGGAGGAGCCTGTGCCGGCAAATGGGTATCAGGAGCCTCAGGAAGTGCAGGAATTCGC 2623
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCAGGAGGAGCCTGTGCCGGCAAATGGGTATCAGGAGCCTCAGGAAGTGCAGGAATTCGC 720

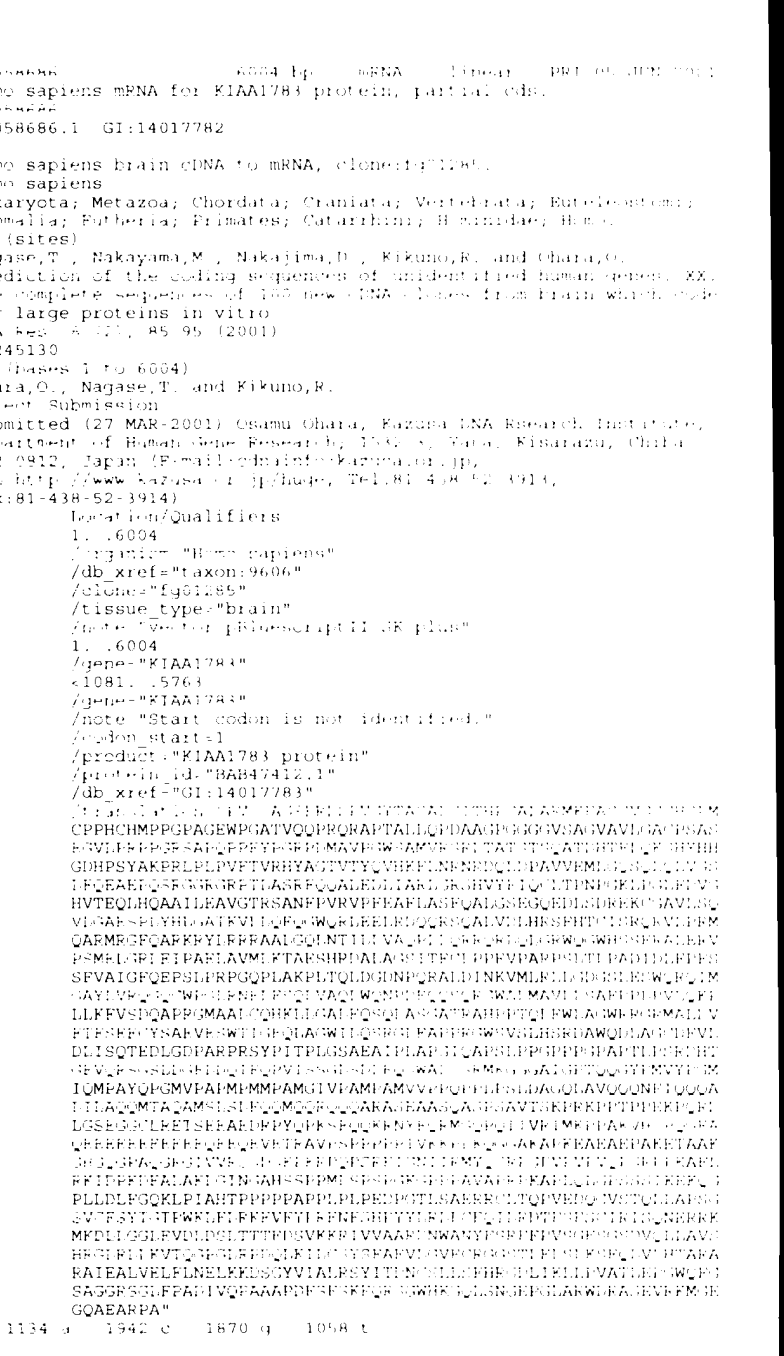
Qy    2624 CCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGCAGGCCCTCAAGAAGTGCCGAATACCT 2683
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGCAGGCCCTCAAGAAGTGCCGAATACCT 780

Qy    2684 CAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCCGCCTGGCACTGGAGAAC 2742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCCGCCTGGCACTGGAGAAC 840

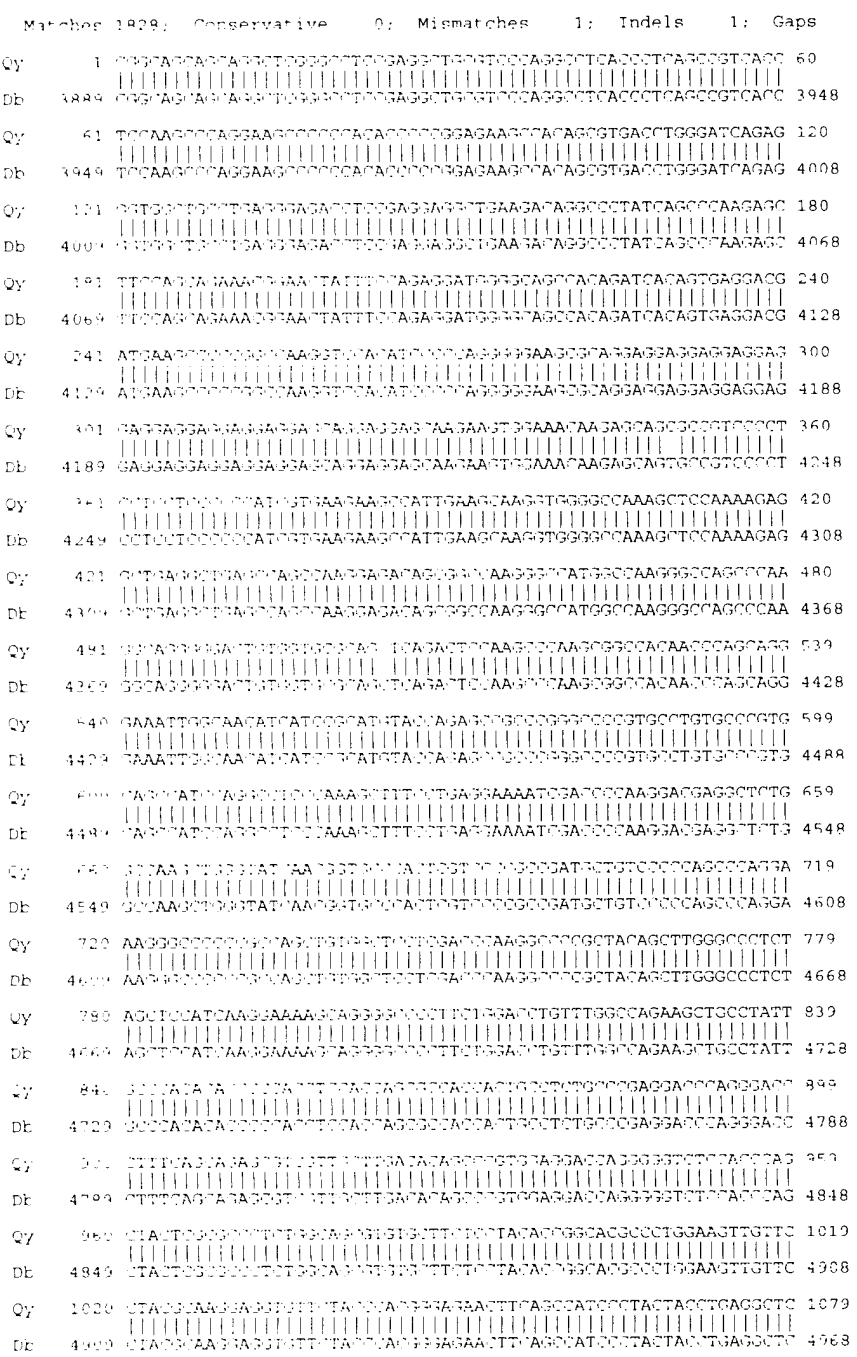
Qy    2743 CGCAGTGCACTTCGATAACTCCACCTACATCAGCAAGCGACTACAGCGAGGTGCTGTGGGA 2802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

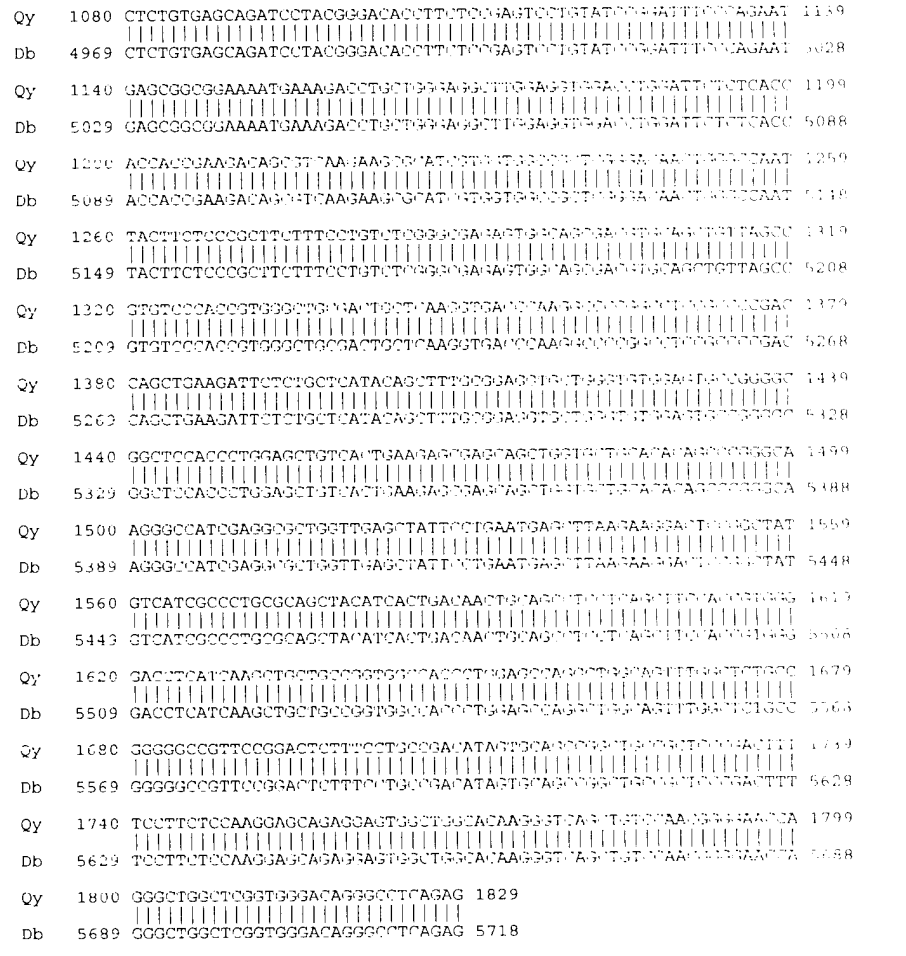


RESULT 7
AB058686



Query Match 48.1%; Score 1816.4; DB 9; Length 6004;
Best Local Similarity 99.9%; Pred. No. 0;





```
RESULT 8
AK097896
LOCUS       AK097896                2372 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ40577 fis, clone THYMU2007467, weakly similar
            to Mus musculus unconventional myosin-15 mRNA.
ACCESSION   AK097896
VERSION     AK097896.1  GI:21757796
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens thymus cDNA to mRNA, clone lib:THYMU2
            clone:THYMU2007467.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
            Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
            Yamashita,H., Matsuo,K., Nakamura,Y., Sekino,M., Kikuchi,N.,
```

```
            Kanda,K., Wajatsuma,M., Murakawa,F., Kanehori,K., Sugiyama,A.,
            Kawakami,B., Ohgiya,Y., Sugino,S., Nagahari,F., Maruno,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2372)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438 52-3975, Fax:81-438 52-3986)
COMMENT     NEDO human cDNA sequencing project (Mot) supported by Ministry of
            Economy, Trade and Industry of Japan. cDNA full insert sequencing:
            Research Association for Biotechnology (RAB), DNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="THYMU2004467"
                     /tissue_type="shymus"
                     /clone_lib="THYMU2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      480 a    780 c    715 g    291 t
ORIGIN
```

```
  Query Match             47.8%;  Score 1808.4;  DB 9;  Length 2372;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1823;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       1 CTGCAGCAGCAGGCTCTGGGCCTCCGAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 CTGCAGCAGCAGGCTCTGGGCCTCCGAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACC 357

Qy      61 TCTAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 TCTAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 417

Qy     121 GCTGGCTGCTCTGAGGGAGACCTGCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 GCTGGCTGCTCTGAGGGAGACCTGCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 477

Qy     181 TTTAGAGAAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 TTTAGAGAAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 537

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCCAGGAGGAGGAGGAGGAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     538 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCCAGGAGGAGGAGGAGGAG 597

Qy     301 GAAGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 GAAGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGTGCCGTCCCCT 657

Qy     361 CCTCCTCCCCCCATCCTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     658 CCTCCTCCCCCCATCCTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 717

Qy     421 CCTCAGCCTCAGCCCAGCCAAGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     718 CCTCAGCCTCAGCCCAGCCAAGAGACAGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAA 777

Qy     481 GCCAGGGGGACTGTGGTGCCGCAG-TCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGG 539
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     778 GCCAGGGGGACTGTGGTGCCGCAGCTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGG 837

Qy     540 GAAATTGGCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCGTG 599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     838 GAAATTGGCAACATCATCCGCATGTACCAGAGCCGCCCGGGCCCCGTGCCTGTGCCCGTG 897

Qy     600 CAGCCATCCAGGGCTCCCAAAGCTTCCTGAGCAAAATCAGCCCAAGCAGGCCTCTG 659
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     898 CAGCCATCCAGGGCTCCCAAAGCTTCCTGAGCAAAATCAGCCCAAGCAGGCCTCTG 957

Qy     660 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCGTGCCAGTGTCTCCCAGCCAGAA 719
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     958 GCCAAGCTGGGTATCAACGGTGCCCACTCGTCCGTGCCAGTGTCTCCCAGCCAGAA 1017

Qy     720 AAGGGCCCGCCTGCCAGTCTGGCTCCTGGACCCAAAGCCCCCTACAGCTTGGCCCTCT 779
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1018 AAGGGCCCGCCTGCCAGTCTGGCTCCTGGACCCAAAGCCCCCTACAGCTTGGCCCTCT 1077

Qy     780 AGTTCCATCAAGGAAAAGCAGGGGCCCCTCTGGACTCTGATTGGCAGAACTGCTATT 839
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1078 AGCTCCATCAAGGAAAAGCAGGGGCCCCTCTGGACTCTGATTGGCAGAACTGCTATT 1137

Qy     840 GCCCACACACCCCACCTCCACCAGCCCACCACTGCCTCTCCCAGAGACCCAGGACC 899
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1138 GCCCACACACCCCACCTCCACCAGCCCACCACTGCCTCTCCCAGAGACCCAGGACC 1197

Qy     900 CTTTCAGCAGAGGGTCGTTGCTTGACACAGCCCGTGAGAGCCAAGGGTCTCCACCCAG 959
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 CTTTCAGCAGAGGGTCGTTGCTTGACACAGCCCGTGGAGAGCCAAGGGTCTCCACCCAG 1217

Qy     960 CTACTGGGGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACCCCTTGGAAGTTGTTC 1019
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 CTACTGGGGCCCTCTGGCAGCGTGTGCTTCTCCTACACCGGCACCCCTTGGAAGTTGTTC 1277

Qy    1020 CTACGCAAGGAGGTGTTCTACCCACAGGAGAACTTCACCATCCTACTATCTGGCCTTC 1079
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 CTACGCAAGGAGGTGTTCTACCCACAGGAGAACTTCACCATCCTACTATCTGGCCTTC 1337

Qy    1080 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTATCCCGHATTTCCCAAAT 1139
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1338 CTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAGTCCTGTATCCCGAATTCCCAAAT 1397

Qy    1140 GAGCGGGGGAAAATGAAAGACCTCCTGGAGGCTTCGACCTTCGATTTCTTCAAT 1199
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1398 GAGCGGGCGGAAAATGAAAGACCTCCTGGAGGCTTCGACCTTCGATTTCTTCAAT 1457

Qy    1200 ACCACCGAAGACAGGGTCAAGAAGCGCATCGTGGTGGCCTTCTCGAACAATCTGGGCAAT 1259
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1458 ACCACCGAAGACAGGGTCAAGAAGCGCATCGTGGTGGCCTTCTCGAACAATCTGGGCAAT 1517

Qy    1260 TACTTCTCCCCGCTTCTTTCCTGTCTGGGCGGGAGTGGCAGCAGCTGCTTACCC 1319
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1518 TACTTCTCCCCGCTTCTTTCCTGTCTGGGCGGGAGTGGCAGCAGCTGCTTACCC 1577

Qy    1320 GTGTCCCACCGTGGGCTGGGACTGCTCAAGGTGACCAAGGCCTGGCCTCTGCCCGAC 1379
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1578 GTGTCCCACCGTGGGCTGGGACTGCTCAAGGTGACCAAGGCCTGGCCTCTGCCCGAC 1637

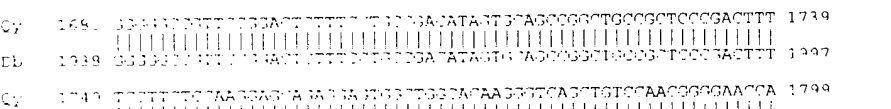
Qy    1380 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGTGGAATGCCGGGC 1439
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1638 CAGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGGTGTGGAATGCCGGGC 1697

Qy    1440 GGCTCCACCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTGACAGCCGGGCA 1499
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1698 GGCTCCACCCTGGAGCTGTCACTGAAGAGCGAGCAGCTGGTGCTGACAGCCGGGCA 1757

Qy    1500 AGGGGCCATCGAGGCGCTGGTTGAGCTATTCCTGAACGAGCCTAAGAAGACTCCGCCTAT 1559
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1758 AGGGGCCATCGAGGCGCTGGTTGAGCTATTCCTGAACGAGCCTAAGAAGACTCCGCCTAT 1817

Qy    1560 GTCATCGGCCCTGCGGCAGCTACATCACTGACAACTGCAGCCTCCCTACCGGGG 1619
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1818 GTCATCGGCCCTGCGGCAGCTACATCACTGACAACTGCAGCCTCCCTACCGGGG 1877

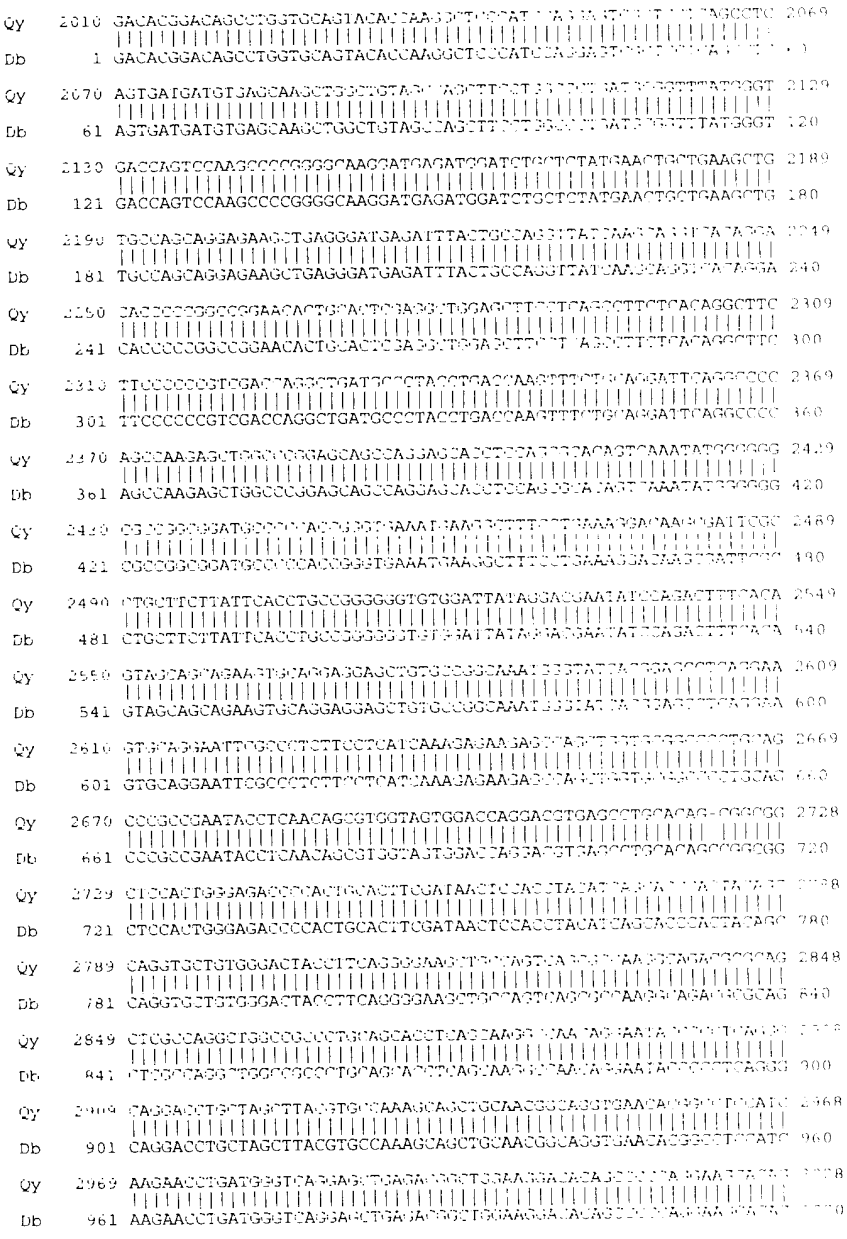
Qy    1620 GACCTCATCAAGCTGCTGCCGGTGGCCACCCTGGAGCCAGGCTCCTCCTGGCCTCCT 1679
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1878 GACCTCATCAAGCTGCTGCCGGTGGCCACCCTGGAGCCAGGCTCCTCCTGGCCTCCT 1937
```

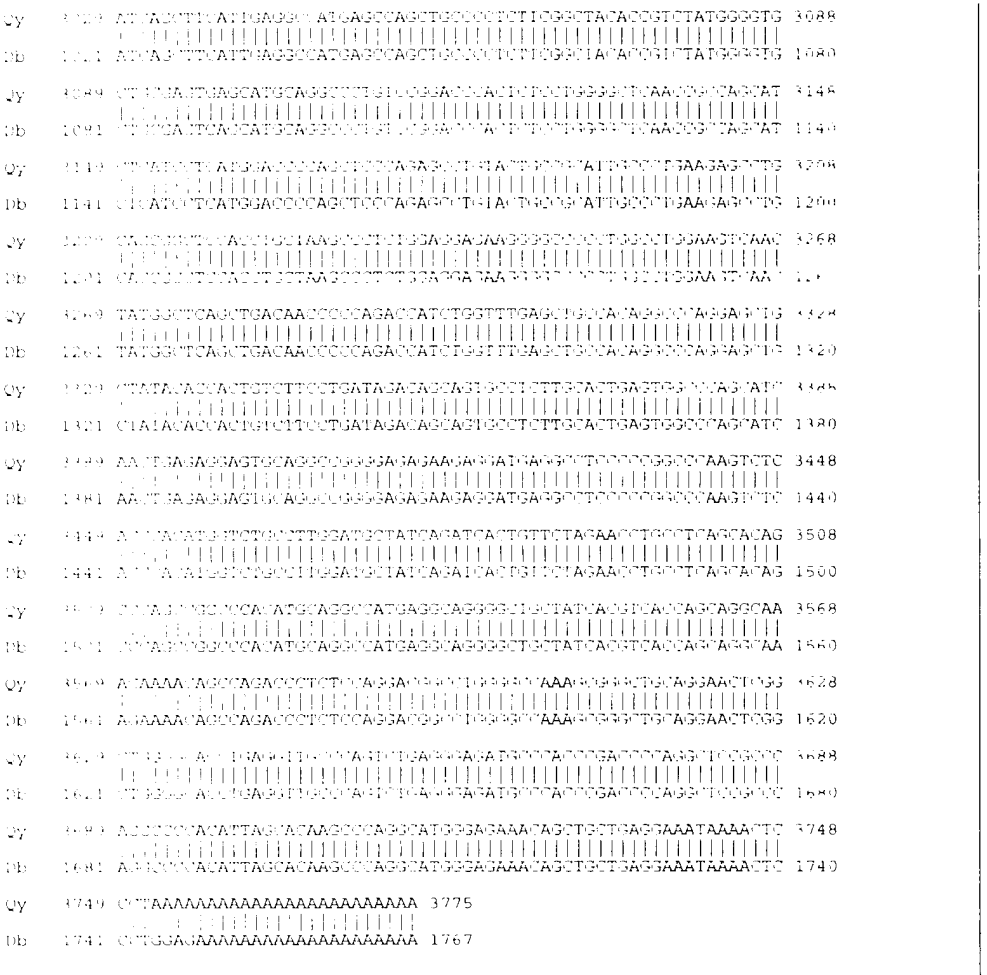


RESULT 9
HSM803107
LOCUS HSM803107 1767 bp mRNA linear PRI 20-MAR-2002
DEFINITION H. sapiens mRNA; cDNA DKFZp667C0210 (from clone DKFZp667C0210)
ACCESSION AL713730
VERSION AL713730.1 GI:19584466
KEYWORDS .
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1767)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp667C0210) is available at the RZPD in Berlin. Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY, Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.

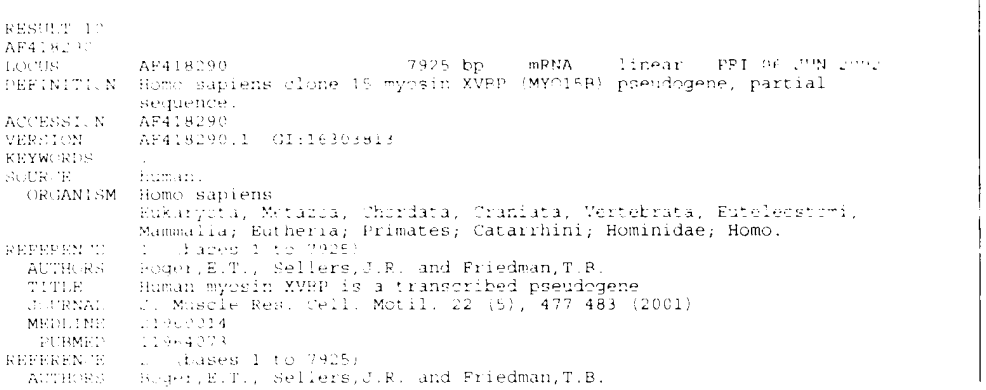
FEATURES	Location/Qualifiers
source	1..1767
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="17"
	/clone="DKFZp667C0210"
	/tissue_type="lymph node"
	/clone_lib="667 (synonym: hlnc2), Vector pSport1, host DH10B, sites: NotI + SalI"
	/dev_stage="adult"
gene	1..1767
	/gene="DKFZp667C0210"
CDS	1..1296
	/gene="DKFZp667C0210"
	/note="similarity to unconventional myosin 15 (Homo sapiens)"
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="CAD28520.1"
	/db_xref="GI:19584467"
	/translation="DTDSLVQYTKAPIQESLLSLSDDVSKLAVASFLALMRFMGDQSK PRGKDEMDLLYELLKLCQQEKLRDEIYCQVIKQVTGHPRPEHCTPGWSFLSLLTGFFP PSTFLMPYLTKFLQDSGPSQFLARSSQEHLQPTVKYGGPRPMPPPGEMKAPLVGQAIP LLLIHLPGGVDYRTNIQTFTVAAEVQEELCRQMGITEPQEVQEFALFLIKEKSQLVRP LQPAEYLNSVVVDQDVSLHSRRLHWETPLHFDNSTYISTHYSQVLWDYLQGKLPVSAK ADAQLARLAALQHLSKANRNTISGQDLLAYVPKQLQRQVNTAS1KNLMGQELSRLEGH SPQEAQISKIEAMSQLPLFGYTVYGVLPVSMQALSGPTLLGLNRQHLILMDPSSQSLY CRIALKSLQRLHLLSPLEEKGPPGLEVNYGSADNPQTIWFELPQAQELLYTTVFLIDG SASCTEWPSIN"
polyA_signal	1731..1736
	/gene="DKFZp667C0210"
polyA_site	1745
	/gene="DKFZp667C0210"

BASE COUNT 418 a 559 c 503 g 287 t
ORIGIN

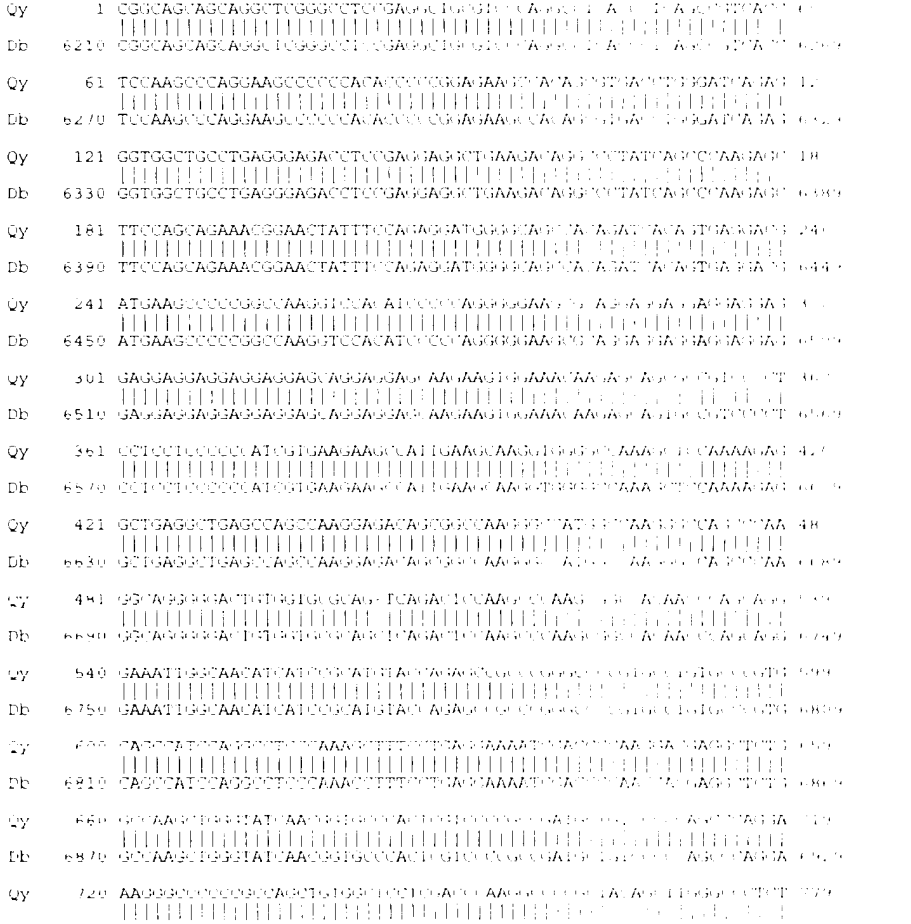


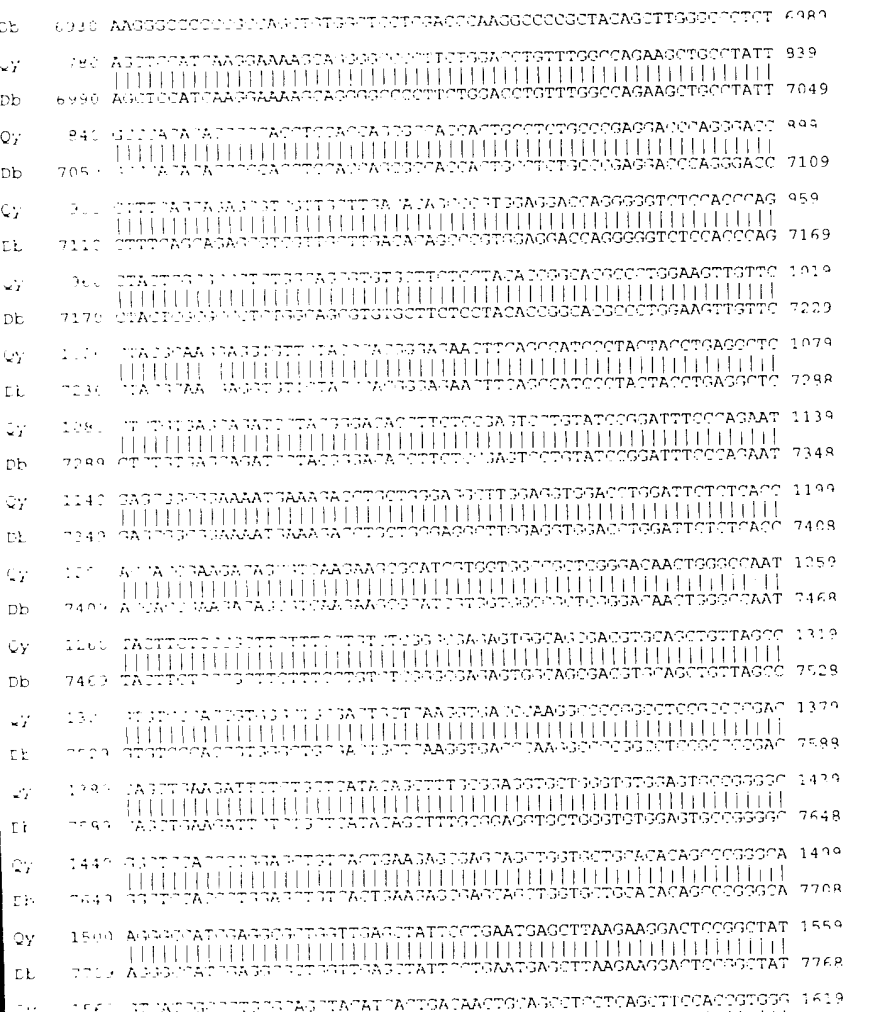


```
RESULT 10
AF418290
LOCUS       AF418290               7925 bp    mRNA    linear   PRI 06-JUN-2001
DEFINITION  Homo sapiens clone 15 myosin XVPP (MYO15B) pseudogene, partial
            sequence.
ACCESSION   AF418290
VERSION     AF418290.1  GI:16306813
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7925)
  AUTHORS   Bogot,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Human myosin XVPP is a transcribed pseudogene
  JOURNAL   J. Muscle Res. Cell. Motil. 22 (5), 477 483 (2001)
  MEDLINE   21566014
    PUBMED  11954023
REFERENCE   2  (bases 1 to 7925)
  AUTHORS   Bogot,E.T., Sellers,J.R. and Friedman,T.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP 2001) Laboratory of Molecular Genetics, Nati
            Institute on Deafness and Other Communication Disorders, 5 Research
            Court, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .7925
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q24 q25"
                     /clone="15"
                     /tissue_type="stomach"
     gene            <1  >7925
                     /gene="MYO15B"
                     /note="XVPP"
                     /pseudo
BASE COUNT     1454 a   2544 c   2691 g   1237 t
ORIGIN
```



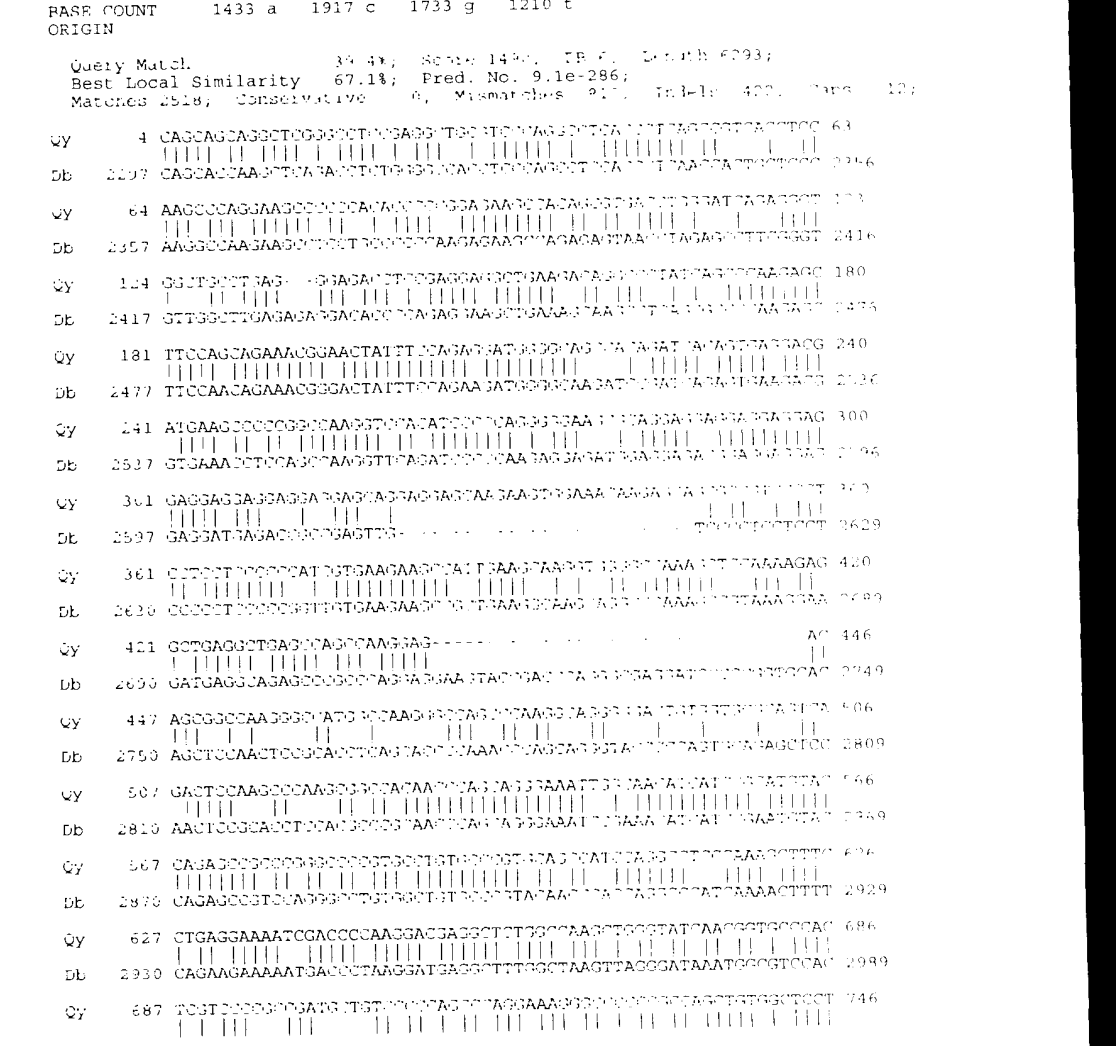
```
  Query Match             44.7%;  Score 1691.4;  DB 9;  Length 7925;
  Best Local Similarity   ++.8%;  Pred. No. 0;
  Matches 1714;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    3;
```

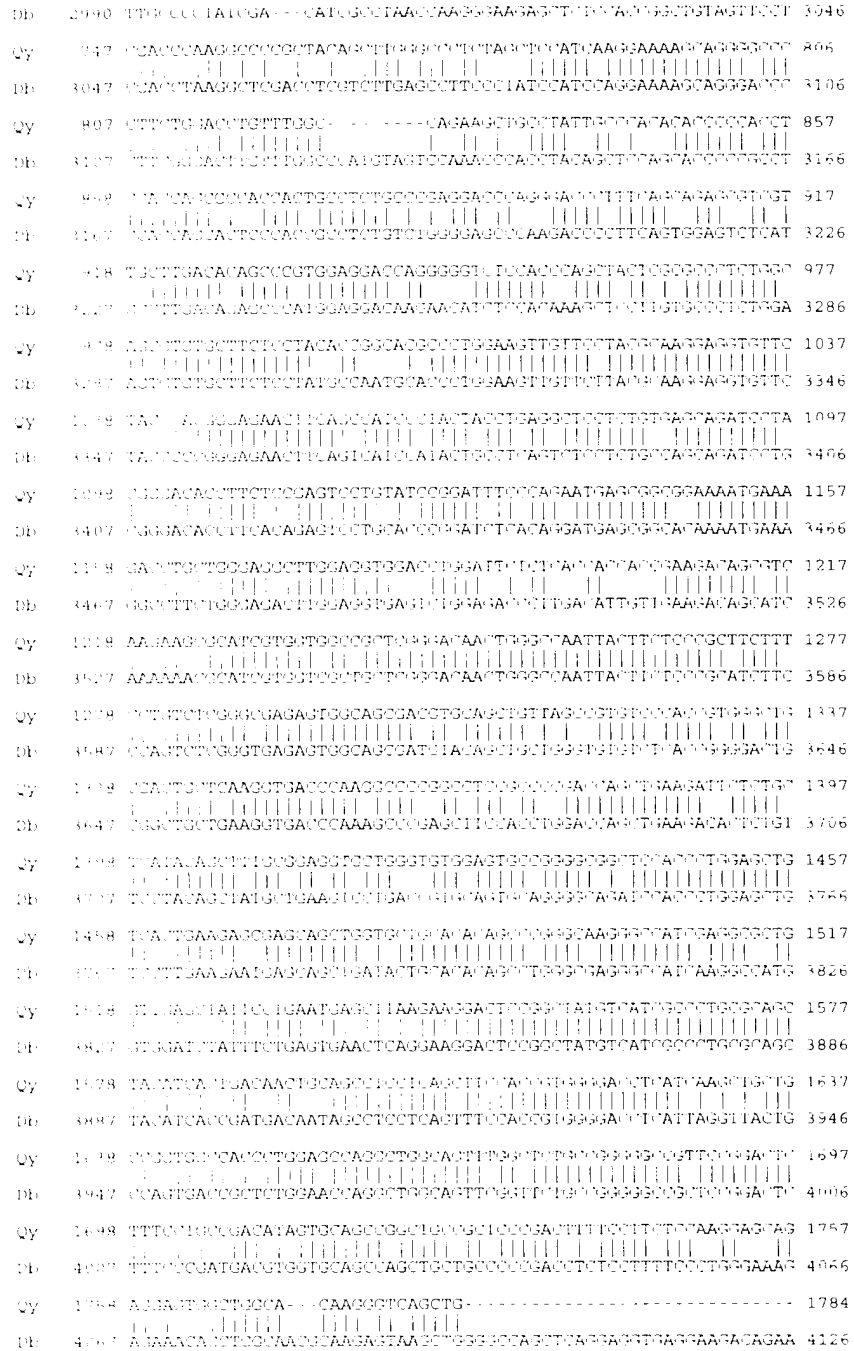


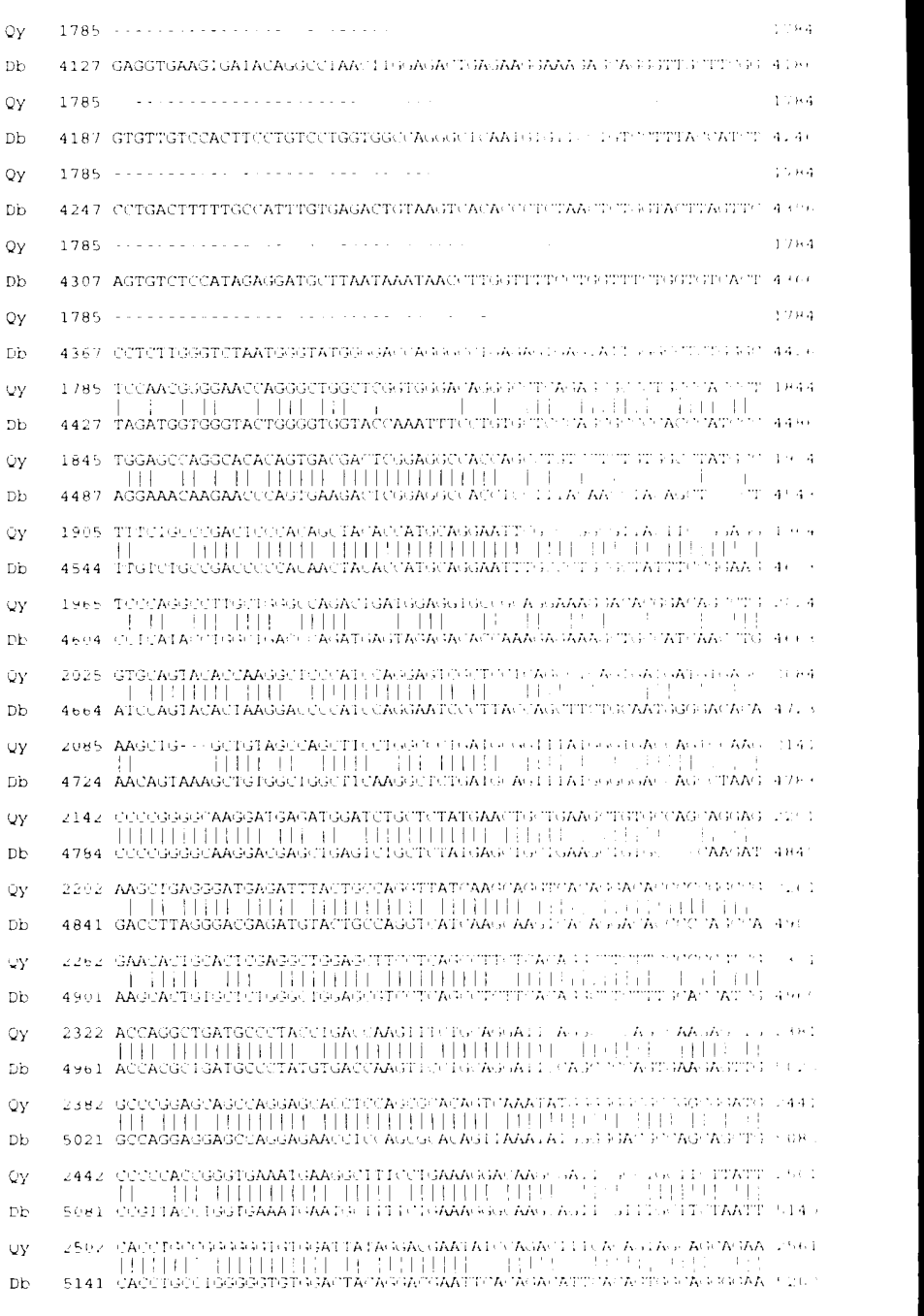


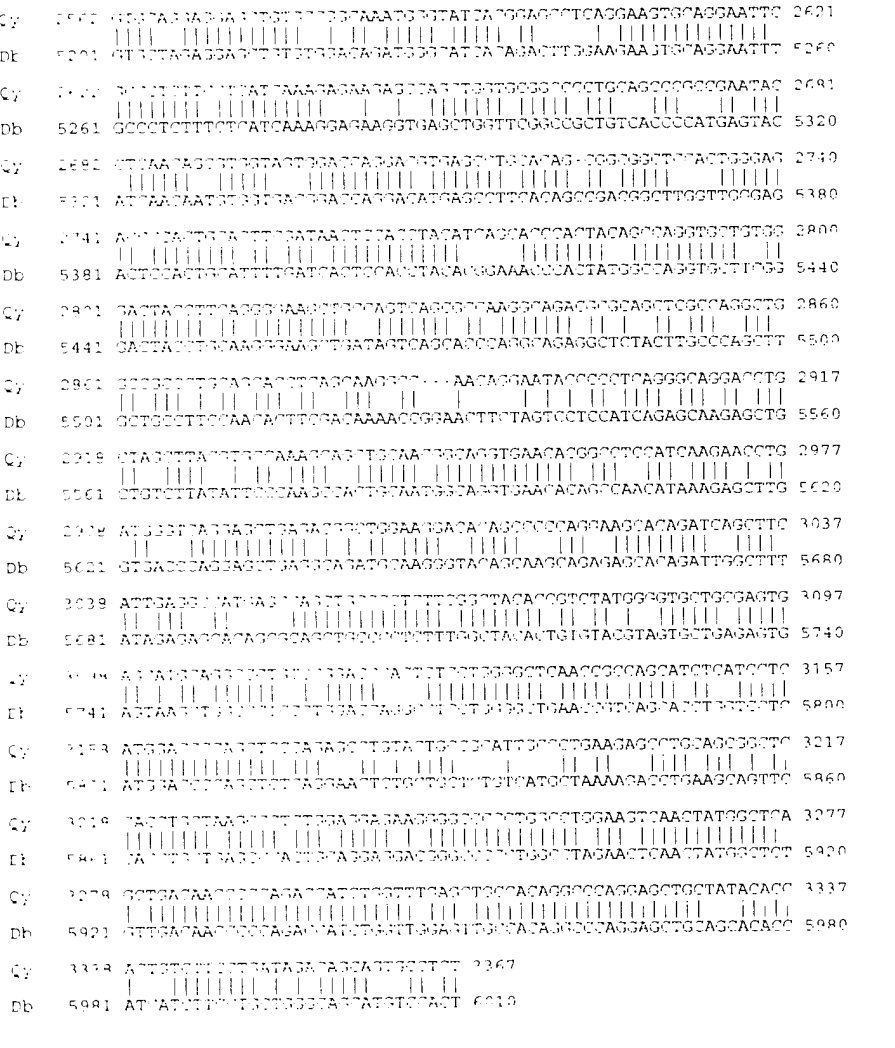
RESULT 11
AX253079
LOCUS AX253079 6293 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 2 from Patent WO0168866.
ACCESSION AX253079
VERSION AX253079.1 GI:15986277

KEYWORDS .
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 6293)
AUTHORS Brooks,A R., Deng,G.G. and Rubanyi,G.M.
TITLE Estrogen-regulated [illegible] protein: compositions and methods of use
JOURNAL Patent WO 0168866 A 2 20 SEP 2001; SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1..6293
/organism="Mus sp."
/db_xref="taxon:10095"
/note="Mouse myosin related protein MYR1 mRNA"
BASE COUNT 1433 a 1917 c 1733 g 1210 t
ORIGIN



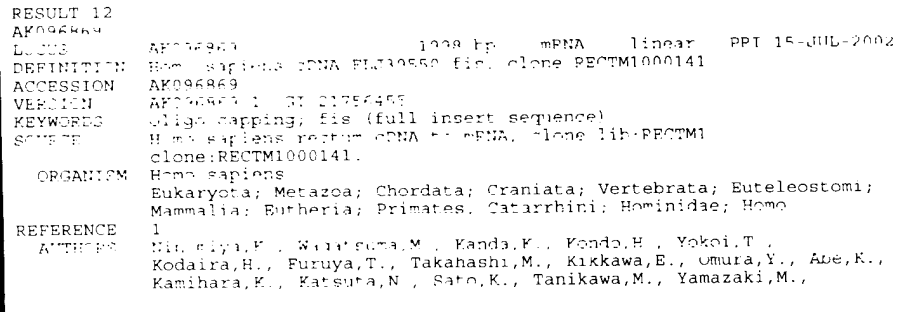


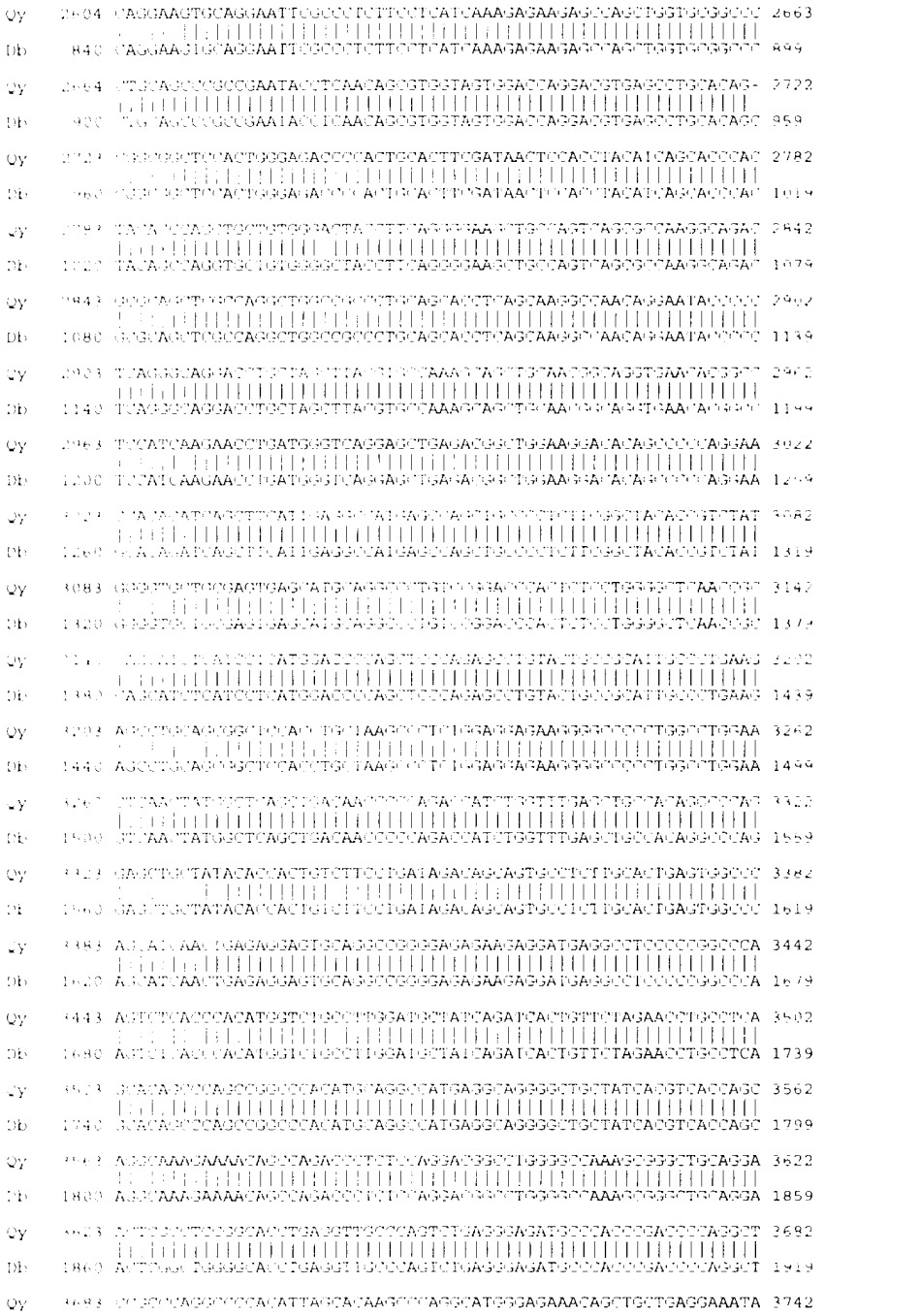


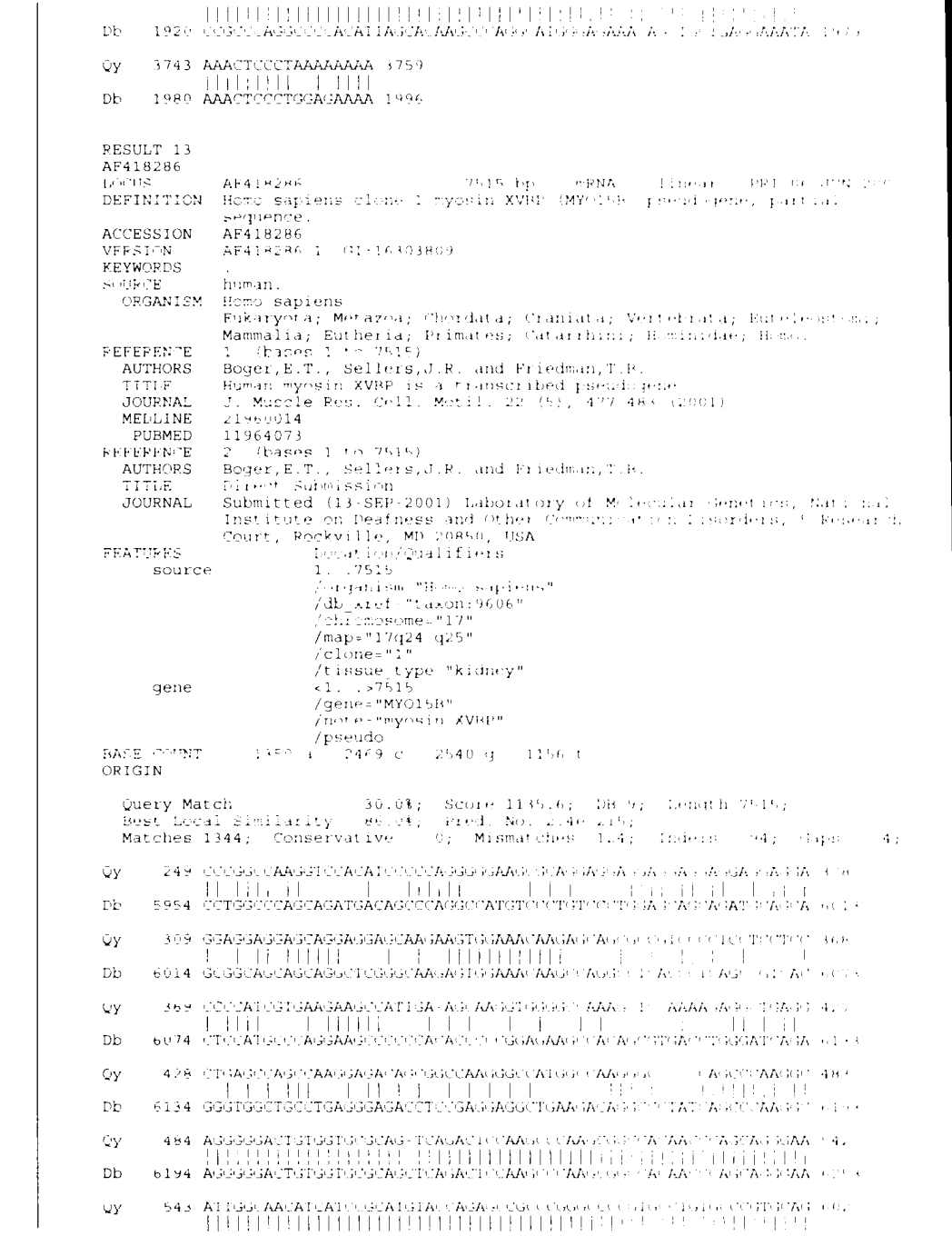


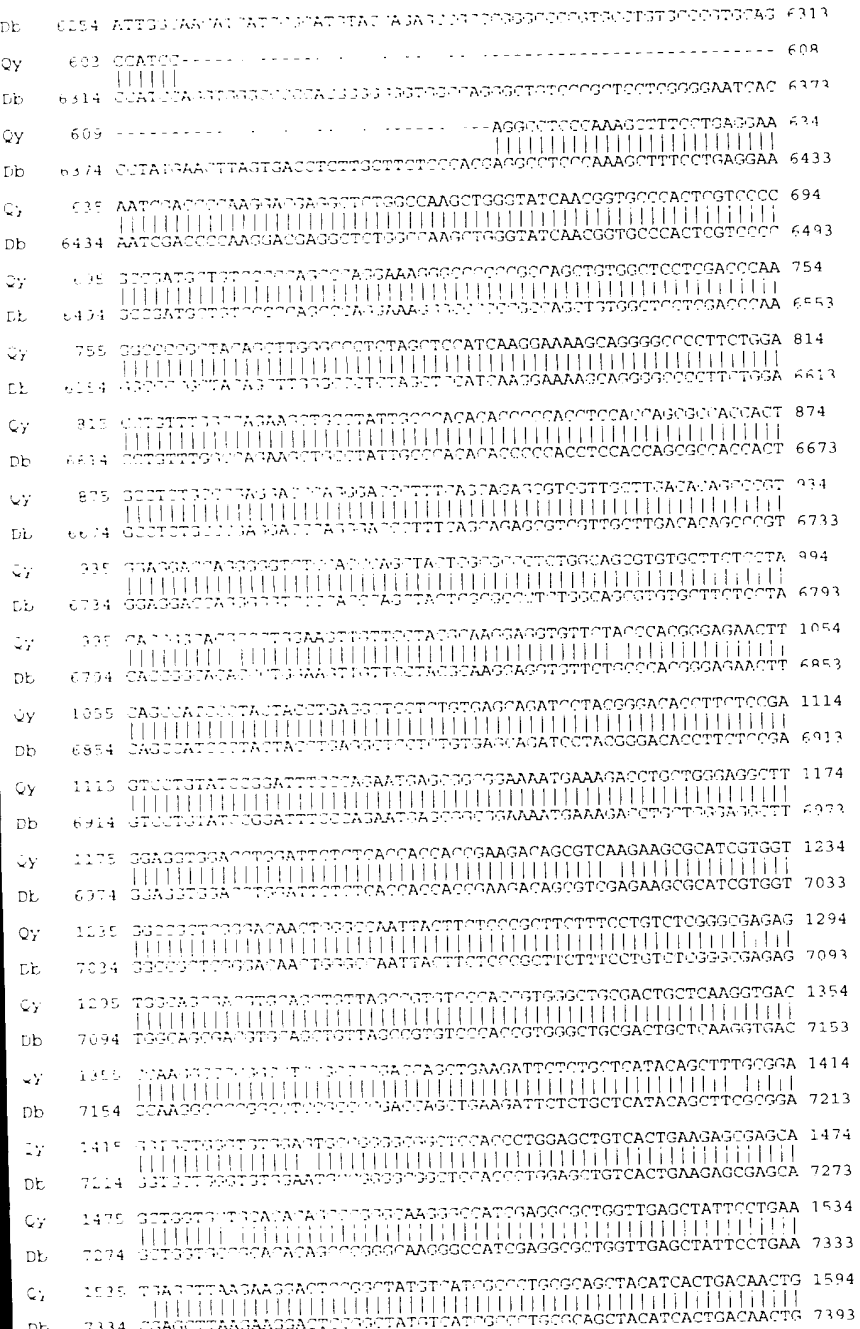
RESULT 12
AK096869
LOCUS AK096869 1998 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ39550 fis, clone RECTM1000141
ACCESSION AK096869
VERSION AK096869.1 GI:21756469
KEYWORDS oligo capping; fis (full insert sequence)
SOURCE Homo sapiens rectum cDNA to mRNA, clone lib:RECTM1
clone:RECTM1000141.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE 1
AUTHORS Nishiyama,E., Watanuma,M., Kanda,K., Kondo,H., Yokoi,T.,
Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hir,Y., Saito,F., Mishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Murakawa,F., Fanenori,F., Takanashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1998)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3927, Fax:81-438-52-3931)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan, cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES Location/Qualifiers
source 1..1998
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RECTM1000141"
/tissue_type="rectum"
/clone_lib="RECTM1"
/note="cloning vector: pME18SFL3"
BASE COUNT 437 a 642 c 596 g 323 t
ORIGIN

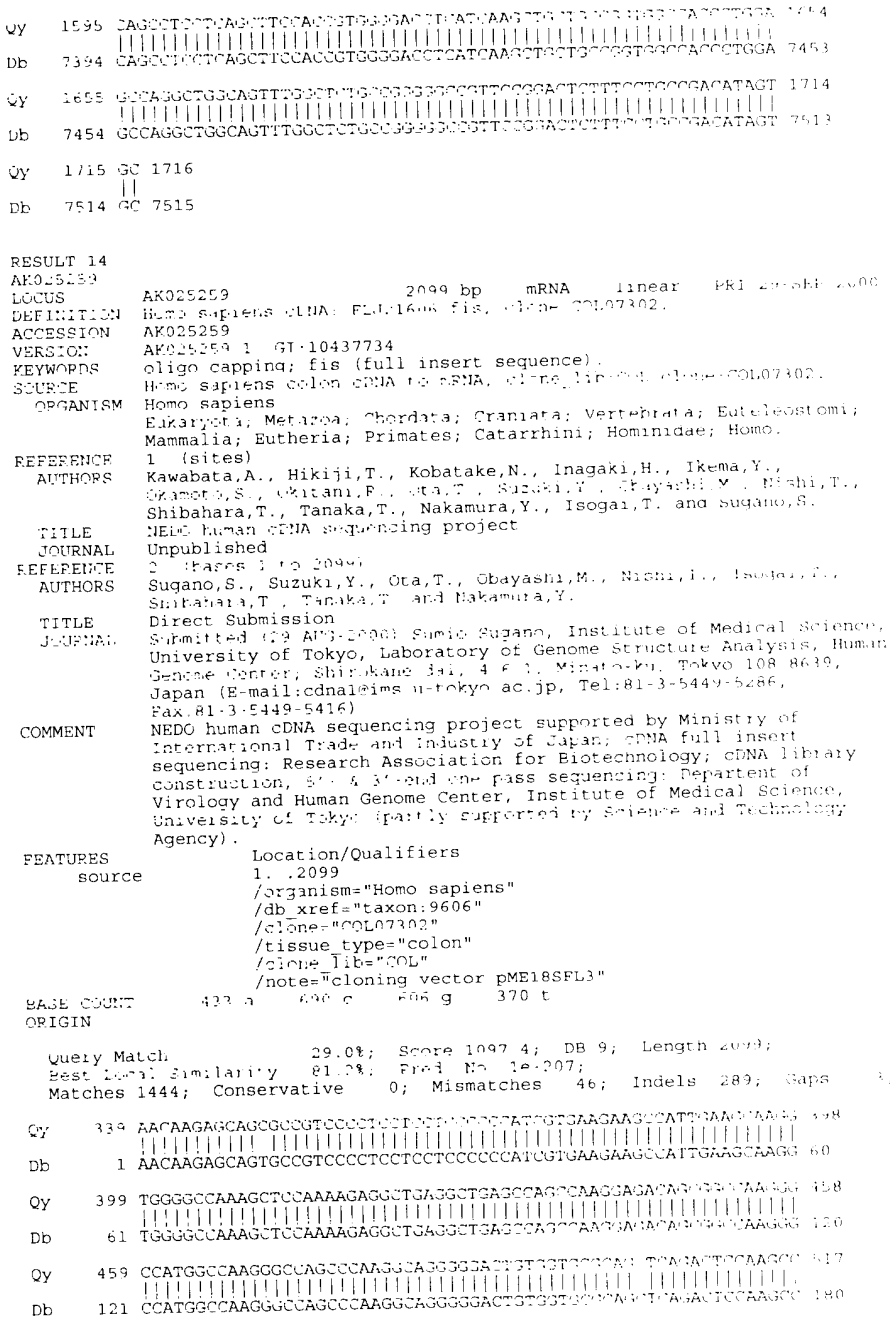
Query Match 36.4%; Score 1376.6; DB 9; Length 1998;
Best Local Similarity 89.3%; Pred. No. 3.4e-263;
Matches 1515; Conservative 0; Mismatches 159; Indels 23; Gaps 2;





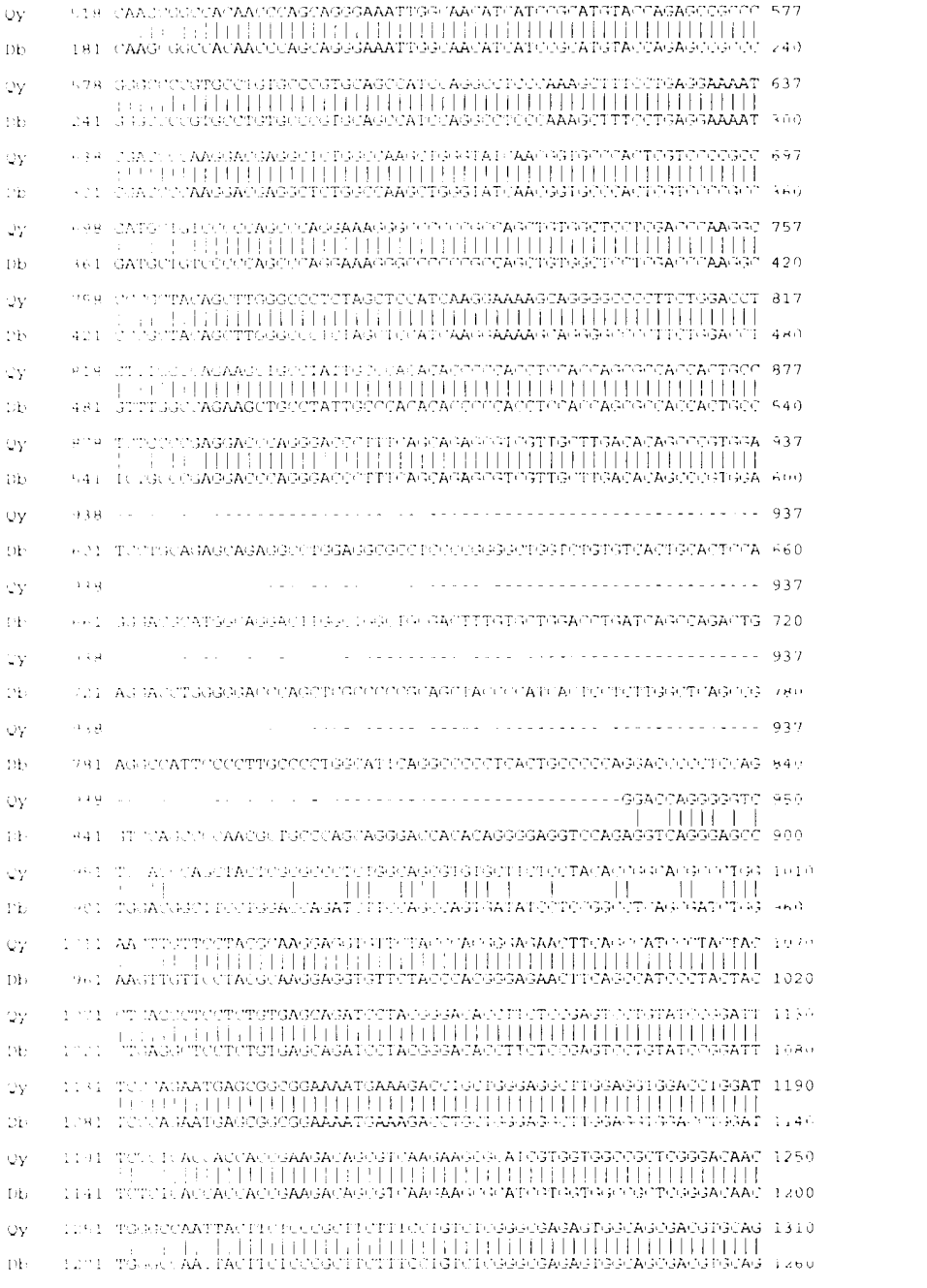


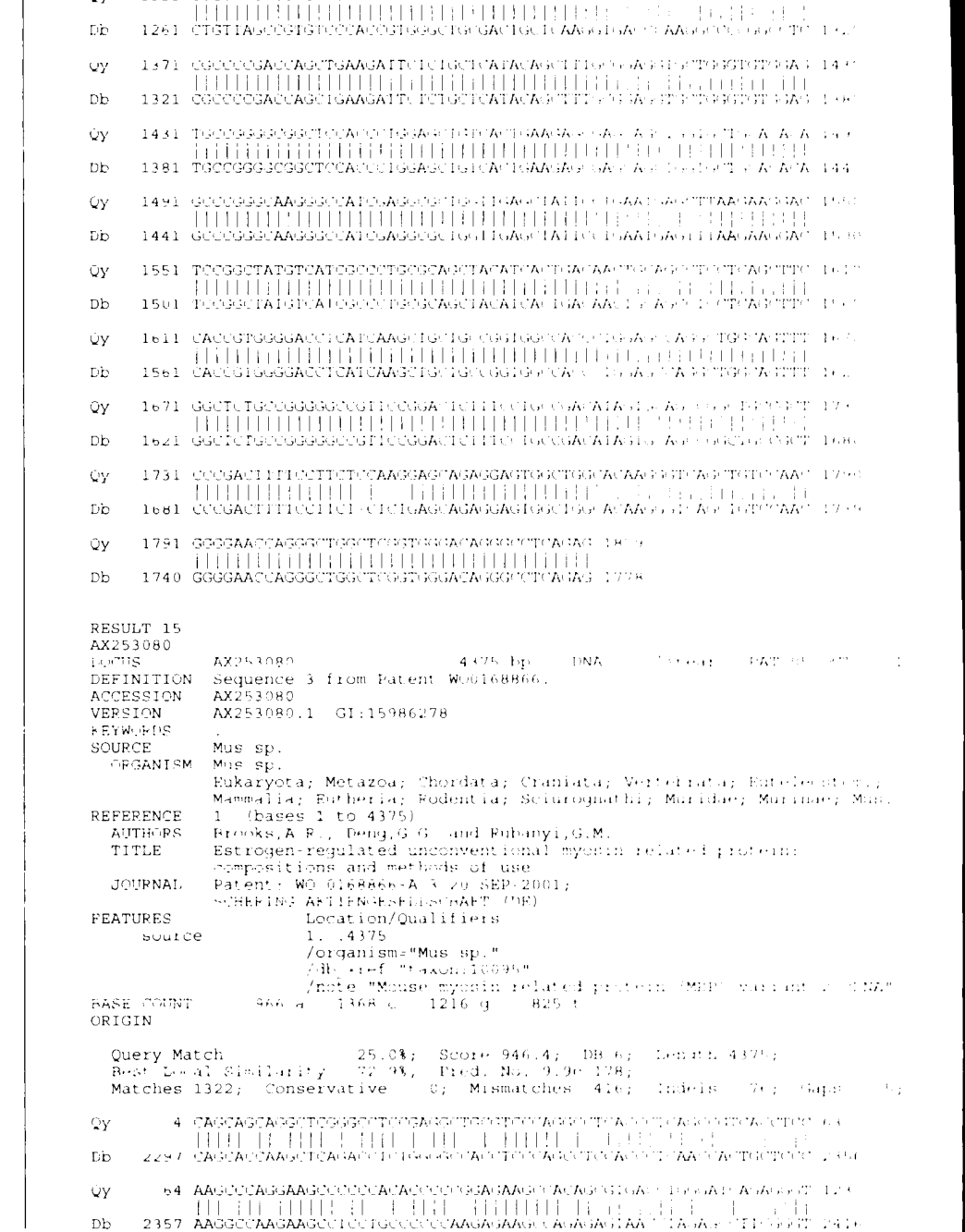




RESULT 14
AK025259
LOCUS AK025259 2099 bp mRNA linear PRI 29-JAN-2003
DEFINITION Homo sapiens cDNA: FLJ21600 fis, clone COL07302.
ACCESSION AK025259
VERSION AK025259.1 GI:10437734
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib=COL, clone=COL07302.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2099)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-APR-2002) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane Dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES Location/Qualifiers
source 1..2099
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL07302"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
BASE COUNT 433 a 696 c 596 g 370 t
ORIGIN

Query Match 29.0%; Score 1097.4; DB 9; Length 2099;
Best Local Similarity 81.0%; Pred. No. 1e-207;
Matches 1444; Conservative 0; Mismatches 46; Indels 289; Gaps 3;



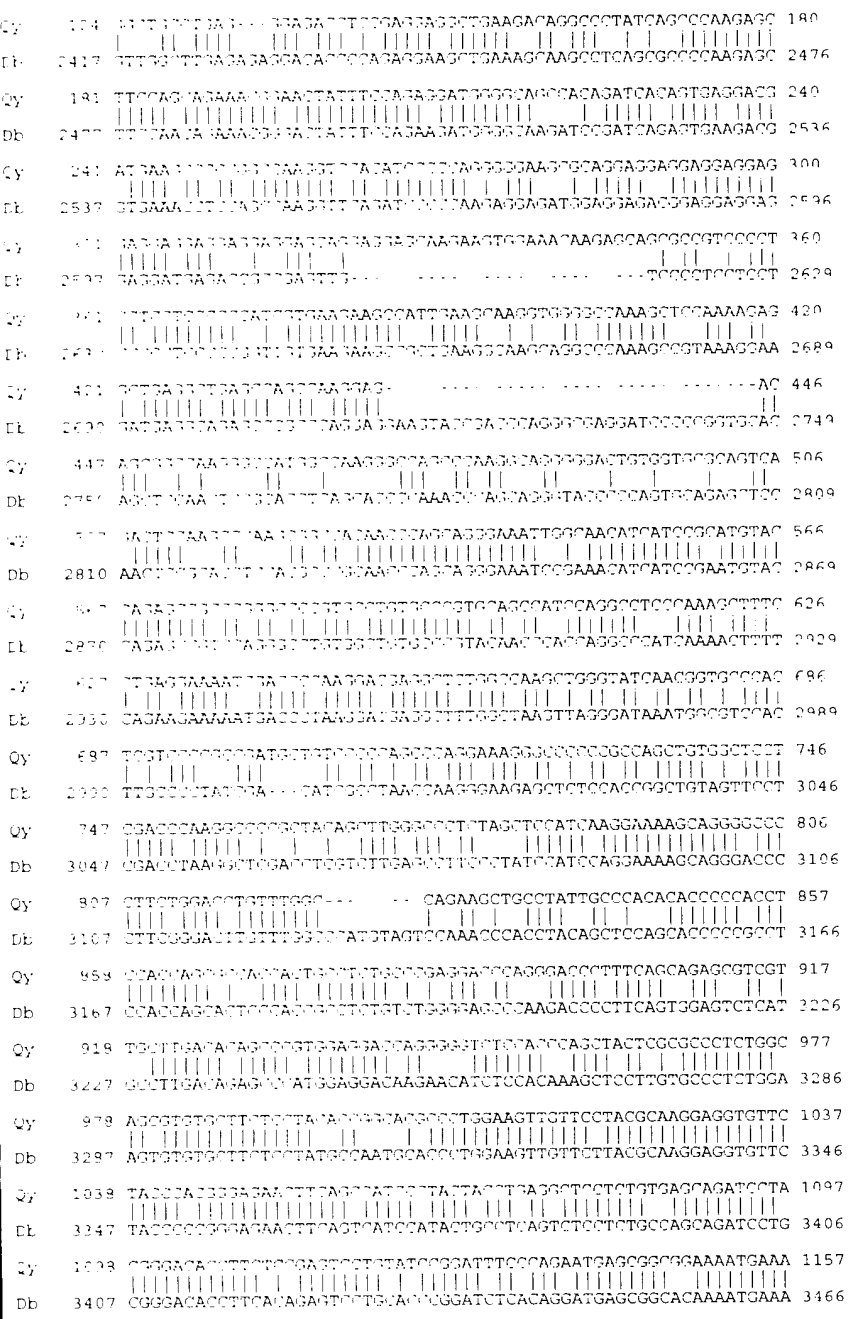


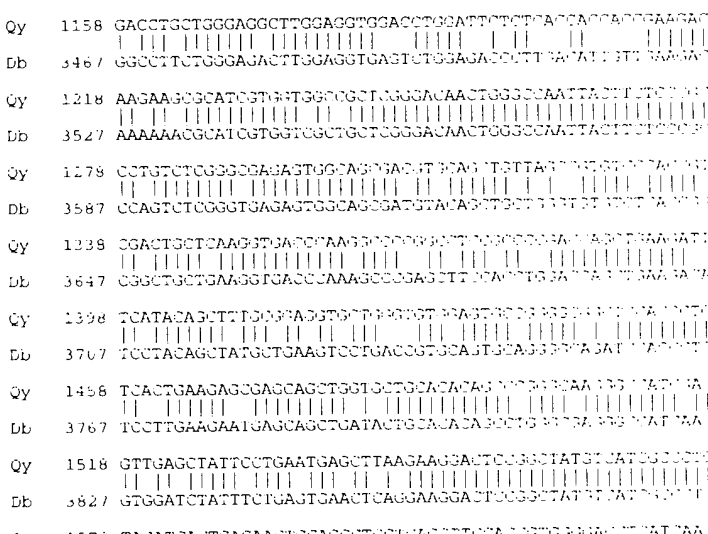
```
RESULT 15
AX253080
LOCUS       AX253080               4375 bp    DNA     linear   PAT 09-JUL-2001
DEFINITION  Sequence 3 from Patent WO0168866.
ACCESSION   AX253080
VERSION     AX253080.1  GI:15986278
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4375)
  AUTHORS   Brooks,A.R., Dong,G.G. and Eubanyi,G.M.
  TITLE     Estrogen-regulated unconventional myosin related proteins:
            compositions and methods of use
  JOURNAL   Patent: WO 0168866-A 3 20-SEP-2001;
            SCHERING AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1..4375
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
                     /note="Mouse myosin related protein 'MRP' variant 2 DNA."
BASE COUNT      966 a   1368 c   1216 g    825 t
ORIGIN

  Query Match             25.0%;  Score 946.4;  DB 6;  Length 4375;
  Best Local Similarity   72.9%;  Pred. No. 9.9e-128;
  Matches 1322;  Conservative    0;  Mismatches  416;  Indels   76;  Gaps    5;

Qy       4 CAGCAGCAGGCTCGGGCCTCCGAGCCTGCCTCCCAGGCTCAACCTCAGCCCCTCACCTCC 63
             |||| || |||| | ||| | |||||| | |||| | |||| | |||| |  |||||
Db    2297 CAGCACCAAGCTCAGACCTCTGGAGCCACCTCCCAGCCTCCAGCCTCAACCTCGCCCCAGC 2356

Qy      64 AAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCCTGACCCGGACCAGGGT 120
           ||| ||| ||||| || | |||| |||||||| || | | | | | | | |||
Db    2357 AAGGCCAAGAAGCCTCCTGCCCCCCAAGAGAAGCCAGAGACTAACCCAGACCTTCCGGT 2416
```





Search completed: January 9, 2003, 20:02:21
Job time : 9369.4 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 02:42:16 ; Search time 657.674 Seconds
(without alignments)
12943.430 Million cell updates/sec

Title: US-09-803-126-7
Perfect score: 3780
Sequence: 1 cggcagcagtcggcttcgggc ... aaaaaaaaaaaaaaaaaaaa 3780

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999150 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001P.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3780	100.0	3780	22	AAS13823	DNA encoding human
2	3352	88.7	4174	22	AAS13821	DNA encoding human
3	1894.4	50.1	1960	24	ABK35127	Human cDNA encodin
4	1490	39.4	6293	22	AAS13819	cDNA #1 encoding m
5	1204.6	31.9	1346	22	AAS33274	DNA encoding human
6	1175	31.1	1476	21	AAZ56716	Human transmembran
7	961.2	25.4	1269	23	ABK43544	DNA encoding novel
8	954.4	25.2	1247	22	AAS27110	cDNA encoding nove
9	954.4	25.2	1247	23	ABK43853	DNA encoding novel
10	946.4	25.0	4375	22	AAS13820	cDNA #2 encoding m
11	442	11.7	9703	22	AAS33480	DNA encoding human
c 12	442	11.7	24757	22	AAS27687	DNA encoding novel
13	442	11.7	24757	22	AAS33481	DNA encoding human
14	323	8.5	1044	22	AAD07791	Human secreted pro
c 15	292	7.7	311	24	ABN95022	Gene #1520 used to
c 16	292	7.7	311	24	ABL62304	Colon adenocarcino
c 17	292	7.7	311	24	ABL64866	Lung cancer relate
c 18	292	7.7	311	24	ABL65551	Lung cancer relate
c 19	174	4.6	526	22	ABA60390	Human foetal liver
c 20	174	4.6	526	22	ABA29621	Probe #7087 for ge
c 21	174	4.6	526	22	AAK08667	Human brain expres
c 22	174	4.6	526	22	AAK34556	Human bone marrow
c 23	174	4.6	526	22	AAI40274	Probe #8960 used t
c 24	174	4.6	526	24	ABS03102	Human genome deriv
c 25	170	4.5	170	22	ABA72929	Human foetal liver
c 26	170	4.5	170	22	ABA38498	Probe #16964 for s
c 27	170	4.5	170	22	AAK21359	Human brain expres
c 28	170	4.5	170	22	AAK47522	Human bone marrow
c 29	170	4.5	170	22	AAI53353	Probe #22039 used
c 30	170	4.5	170	24	ABS21624	Human genome deriv
c 31	137.8	3.6	233	22	ABA74633	Human foetal liver
c 32	137.8	3.6	233	22	ABA39397	Probe #17863 for g
c 33	137.8	3.6	233	22	AAK23104	Human brain expres
c 34	137.8	3.6	233	22	AAK49279	Human bone marrow
c 35	137.8	3.6	233	22	AAI55120	Probe #23866 used
c 36	137.8	3.6	233	24	ABS22813	Human genome deriv
c 37	137.8	3.6	566	22	ABA62139	Human foetal liver
c 38	137.8	3.6	566	22	ABA29547	Probe #8013 for ge
c 39	137.8	3.6	566	22	AAK10458	Human brain expres
c 40	137.8	3.6	566	22	AAK36359	Human bone marrow
c 41	137.8	3.6	566	22	AAI42087	Probe #10773 used
c 42	137.8	3.6	566	24	ABS10413	Human genome deriv
43	70.4	1.9	569	23	AAS76326	DNA encoding novel
44	62.2	1.6	1851	23	ABL04085	Drosophila melano
45	61.4	1.6	693	23	AAS74340	DNA encoding novel

ALIGNMENTS

RESULT 1
AAS13823
ID AAS13823 standard; cDNA; 3780 BP
XX
AC AAS13823;
XX
DT 18-DEC 2001 (first entry)
XX
DE DNA encoding human myosin-related protein, hMRP #2
XX
KW Human; oestrogen regulated unconventional myosin-related protein; MRP;
KW auditory, antiatherosclerotic, osteopathic, neuroprotective; nootropic;
KW antiparkinsonian; cytostatic, oestrogen receptor, atherosclerosis;
KW osteoporosis, breast cancer; cardiovascular disease, deafness; ss;
KW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 561..2911
FT /*tag= a
FT /product= "hMRP"
FT /note= "Oestrogen-regulated unconventional myosin-
FT related protein"
XX
PN WO200168866-A2.
XX
PD 20 SEP 2001.
XX
PF 12 MAR-2001; 2001WO-US08060.
XX
PR 10 MAR-2000; 2000US-188488P.

```
PR   04 MAR 2001; 2001US-0803126.
XX
PA   (GENE ) GENENRING AG.
XX
PI   Brand AP, Peng CG, Rabanyi PM,
XX
DR   WPI; 2001-602747/68.
DR   P-PSDB; AAC07406.
XX
PT   Novel oestrogen-regulated unconventional myosin-related protein and
PT   polynucleotide for modulating effects of oestrogen in mammalian cells to
PT   treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS   Claim 4; Fig 7; 108pp; English.
XX
CC   The invention relates to an isolated oestrogen-regulated unconventional
CC   myosin related protein (MRP). Modulating MRP is useful for modulating the
CC   effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC   mammalian cell. When the level of expression or activity of MRP is
CC   increased, the development of atherosclerosis or osteoporosis is
CC   inhibited and when the level is decreased the development of breast
CC   cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC   useful for detecting the presence of oestrogen signaling in a mammalian
CC   cell, by detecting the expression or activity of MRP or the level of MRP
CC   mRNA in the cell. The presence of oestrogen signaling is used to
CC   determine the responsiveness of the cell to the oestrogen or the tissue-
CC   specific distribution of oestrogen signaling in a mammal. MRP is useful
CC   as a marker gene for the investigation of tissue-specific and oestrogen
CC   receptor-specific ligands such as oestrogen receptor antagonists and
CC   agonists. Modulators, recombinant forms, derivatives, variants or
CC   fragments of MRP are useful for enhancing or inhibiting oestrogen
CC   signaling in cells and for treatment of breast cancer or other cancers
CC   and for preventing or treating wide variety of oestrogen-related
CC   diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC   disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC   prevention and/or treating hearing loss and deafness. MRP polypeptides
CC   are useful for generating monoclonal and polyclonal antibodies useful for
CC   identifying oestrogen responsive cells, particularly cells responsive to
CC   oestrogen through an oestrogen receptor alpha or beta. To identify
CC   alleles, interspecies homologues and polymorphic variants and in
CC   immunoassays to detect MRP. The present sequence represents the coding
CC   sequence of human MRP variant #2 as described in the invention
XX
SQ   Sequence 3780 BP; 838 A; 1228 C; 1107 G; 607 T; 0 other;

  Query Match             100.0%;  Score 3780;  DB 22;  Length 3780;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3780;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TGGCAGCAGAGGCTCGGGGCCTCCGAGGCTGCGTCCACAGCCTCACCCTCACAGCGGTCACC  60
Db       1 TGGCAGCAGAGGCTCGGGGCCTCCGAGGCTGCGTCCACAGCCTCACCCTCACAGCGGTCACC  60

Qy      61 TTCCAAGCCCAGGAAGCCCCCCACACCCCCGAGAAGCCACAGCGGTGACCTGGGATCAGAG  120
Db      61 TTCCAAGCCCAGGAAGCCCCCCACACCCCCGAGAAGCCACAGCGGTGACCTGGGATCAGAG  120

Qy     121 GATGACCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC  180
Db     121 GATGACCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC  180

Qy     181 TTCTAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG  240
Db     181 TTCTAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG  240

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG  300
Db     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG  300

Qy     301 GAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGGCGGTCCCCT  360
Db     301 GAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCCGGTCCCCT  360

Qy     361 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGAAGAGGCGGAAAAAGTCCAAAAGAG  420
Db     361 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGAAGAGGCGGAAAAAGTCCAAAAGAG  420

Qy     421 GCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAGGCCATGGCCAAGGGCAGCCAA  480
Db     421 GCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAGGCCATGGCCAAGGGCAGCCAA  480

Qy     481 GGCAGGGAGACTGTGGTGGGCAGTCAGACTCCAAGCCAAGCAGAAGCAGCAGTAGG  540
Db     481 GGCAGGGAGACTGTGGTGGGCAGTCAGACTCCAAGCCAAGCAGAAGCAGCAGTAGG  540

Qy     541 AAATTGGCAACATCATCCGCATGTACCAGAGCCGCCGCCCCCCCGCCGCCTCCC  600
Db     541 AAATTGGCAACATCATCCGCATGTACCAGAGCCGCCGCCCCCCCGCCGCCTCCC  600

Qy     601 AGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATTGACCCCAAGGACGAGGCTCTGG  660
Db     601 AGCCATCCAGGCCTCCCAAAGCTTTCCTGAGGAAAATTGACCCCAAGGACGAGGCTCTGG  660

Qy     661 CCAAGCTGGGTATCAACGGTGCCACTCCTGCCGAGCGAGCCCCAGCCCAGCAA  720
Db     661 CCAAGCTGGGTATCAACGGTGCCACTCCTGCCGAGCGAGCCCCAGCCCAGCAA  720

Qy     721 AGGGCCCCCCGCCAGCTGTGGCTCCTCAACCAAGCCCCCCCCCCCCCCCCCCCCCCCCTA  780
Db     721 AGGGCCCCCCGCCAGCTGTGGCTCCTCAACCAAGCCCCCCCCCCCCCCCCCCCCCCCCTA  780

Qy     781 GCTCCATCAAGGAAAAGCAAGGGGCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTG  840
Db     781 GCTCCATCAAGGAAAAGCAAGGGGCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTG  840

Qy     841 CCCACACACCCCCACCTCCACCAGGGCCACCACTGCCCTCCCCAGGACCCAGGACCC  900
Db     841 CCCACACACCCCCACCTCCACCAGGGCCACCACTGCCCTCCCCAGGACCCAGGACCC  900

Qy     901 TTTCAGCAGAGGGTCGTGTGCTTGACCACAGCCGTGGAAGACCAAAAGTTCTTCACCAGC  960
Db     901 TTTCAGCAGAGGGTCGTGTGCTTGACCACAGCCGTGGAAGACCAAAAGTTCTTCACCAGC  960

Qy     961 TACTCGCGGCCCTCTGGCAGCGTGTGCTTCTCTACAGCCCCA  1020
Db     961 TACTCGCGGCCCTCTGGCAGCGTGTGCTTCTCTACAGCCCCA  1020

Qy    1021 TACCCAAGGAGGTGTTCTACCCACGGGACAACTTCAGCATCTACTACTGACCTCC  1080
Db    1021 TACCCAAGGAGGTGTTCTACCCACGGGACAACTTCAGCATCTACTACTGACCTCC  1080

Qy    1081 TCTGTGAGCAGATCCTACGGGACACTTCTTCCCAGTCCTCTACCCAGAATC  1140
Db    1081 TCTGTGAGCAGATCCTACGGGACACTTCTTCCCAGTCCTCTACCCAGAATC  1140

Qy    1141 AGCGGCGGAAAATGAAAGACCTGCTGGGAAGGCTTGGAGGTGGACCTACTTCTACTA  1200
Db    1141 AGCGGCGGAAAATGAAAGACCTGCTGGGAAGGCTTGGAGGTGGACCTACTTCTACTA  1200

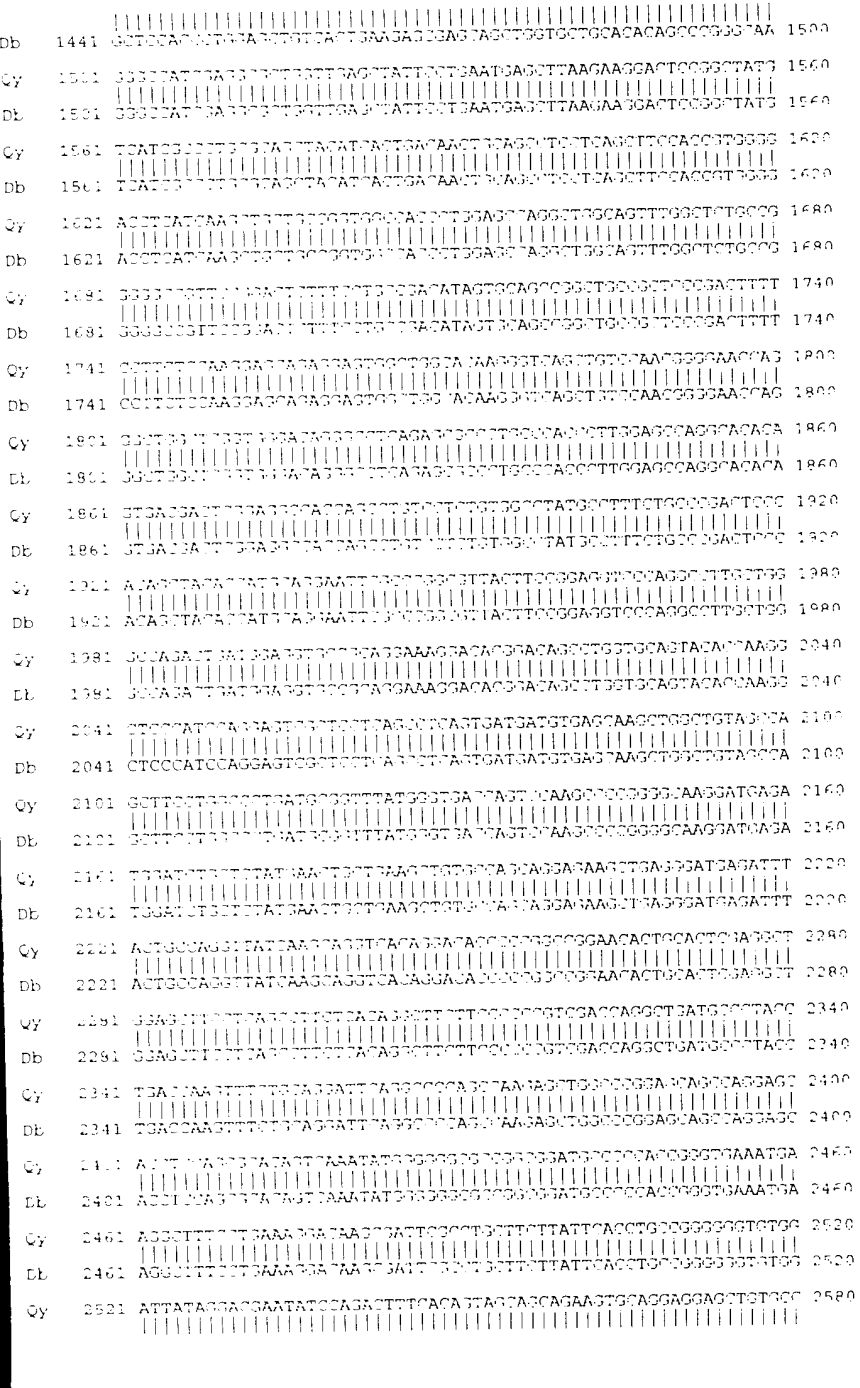
Qy    1201 CCACCGAAGACAGGGTCAAGAAGGCATCGTGGTGGCCGCTCGGGACAACTTGGCCAATT  1260
Db    1201 CCACCGAAGACAGGGTCAAGAAGGCATCGTGGTGGCCGCTCGGGACAACTTGGCCAATT  1260

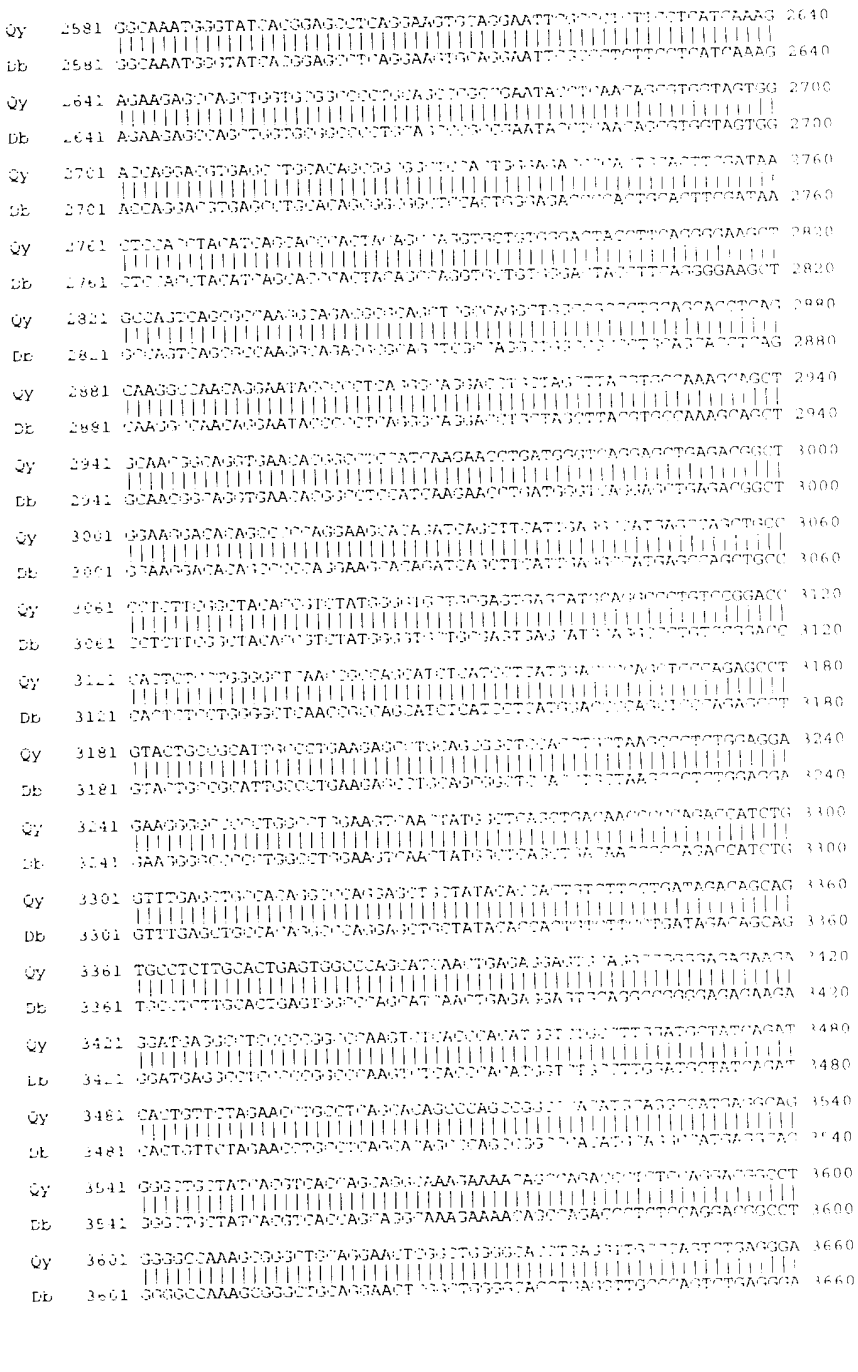
Qy    1261 ACTTCTCCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGCGACGTCGACGCTGTTACGCTG  1320
Db    1261 ACTTCTCCCGCTTCTTTCCTGTCTCGGGCGAGAGTGGCAGCGACGTCGACGCTGTTACGCTG  1320

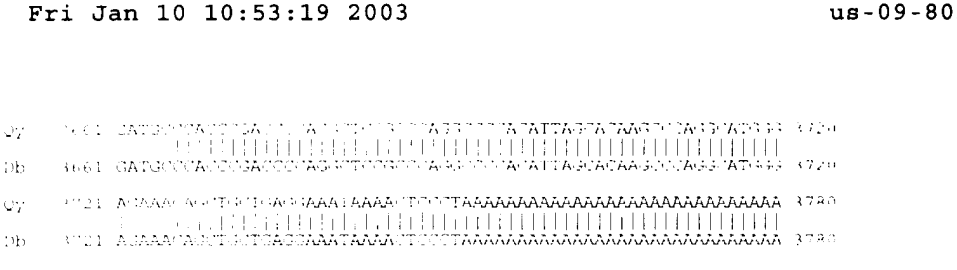
Qy    1321 TGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGGCCCGGCCTCCGCCGAC  1380
Db    1321 TGTCCCACCGTGGGCTGCGACTGCTCAAGGTGACCCAAGGCCCGGCCTCCGCCGAC  1380

Qy    1381 AGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGCTGTGAACTTCCGGGTG  1440
Db    1381 AGCTGAAGATTCTCTGCTCATACAGCTTTGCGGAGGTGCTGGCTGTGAACTTCCGGGTG  1440

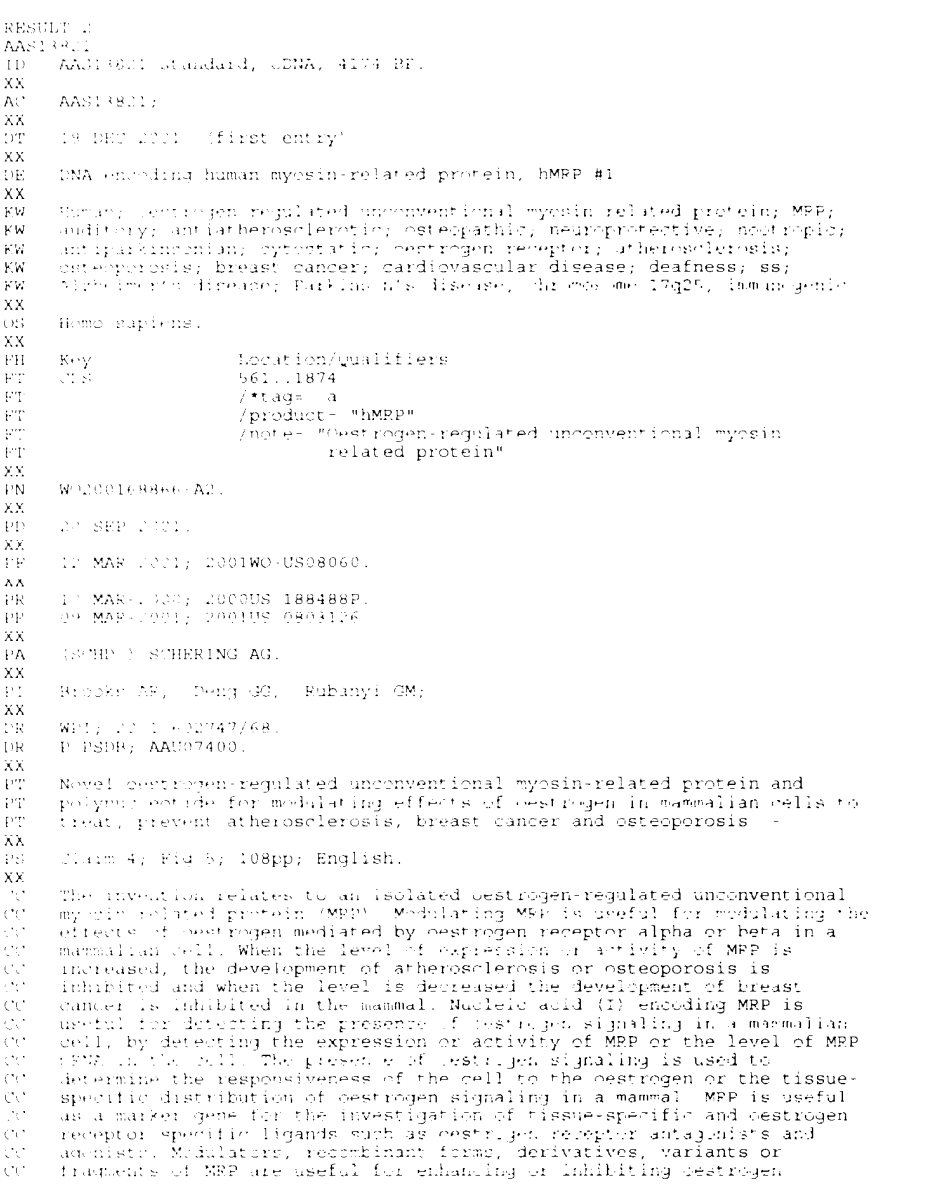
Qy    1441 GCTCCACCCTGGAGCTGTCACTGAAGAGCGCAGCAGCTGGTGCTGCACAACAACTAA  1500
```



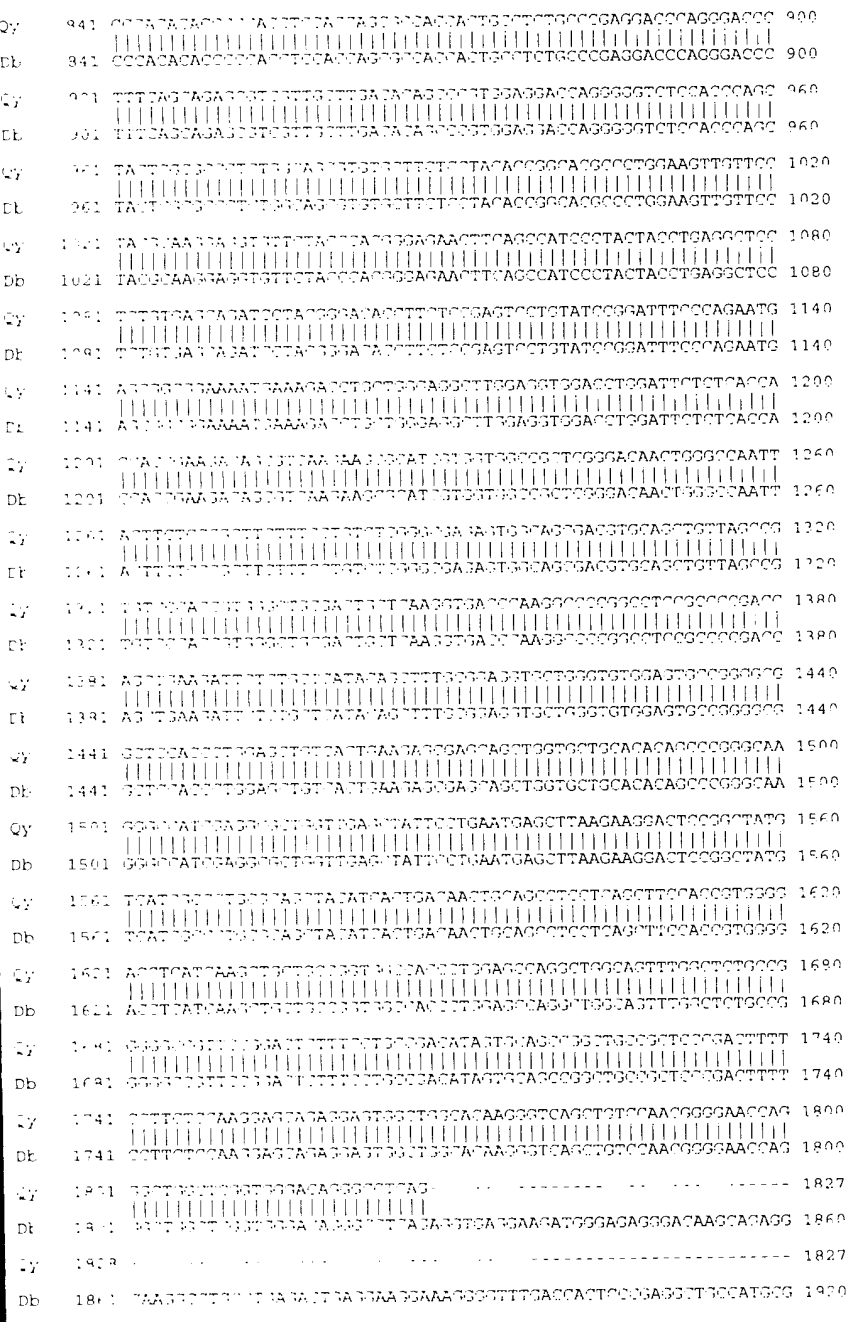


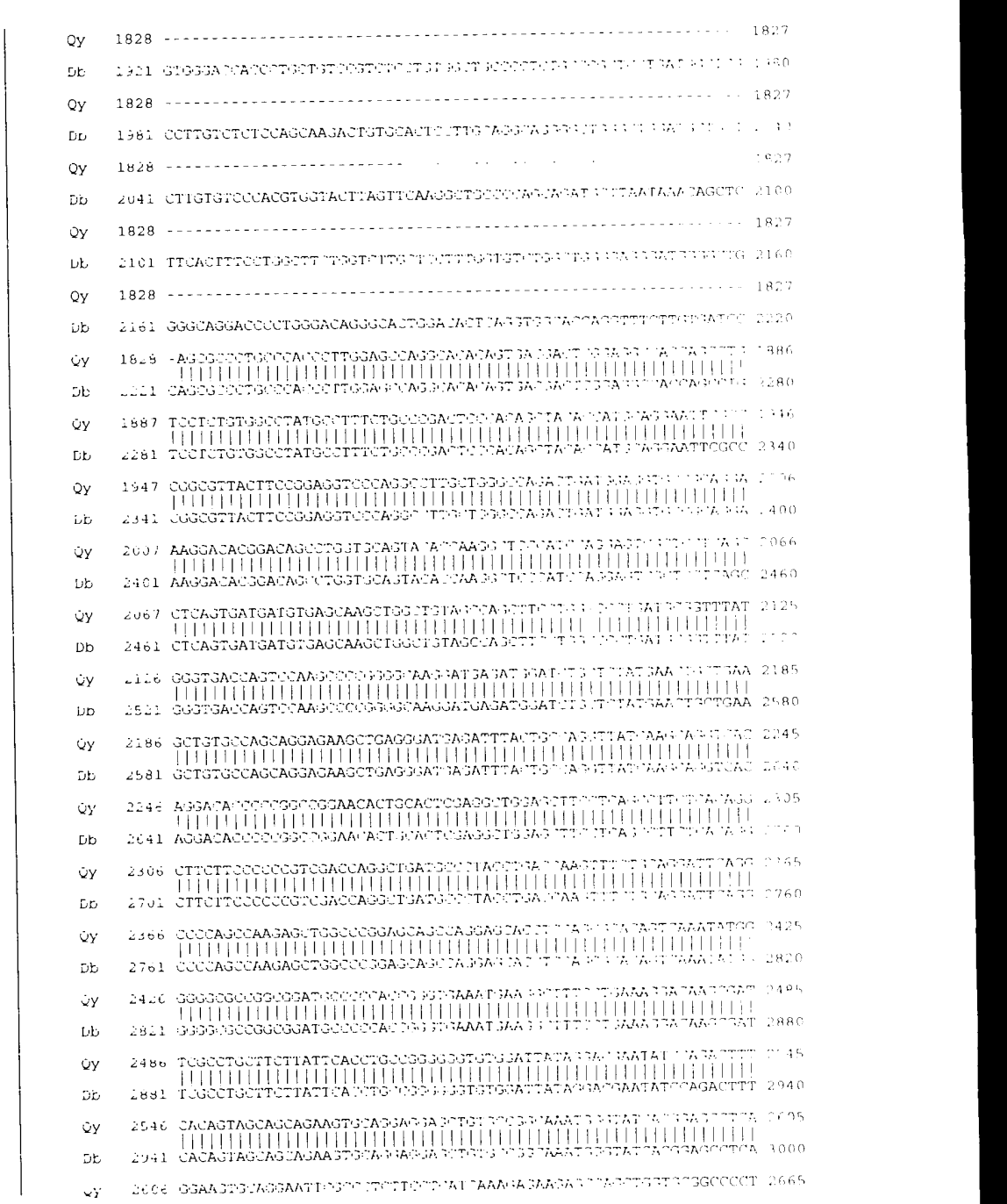


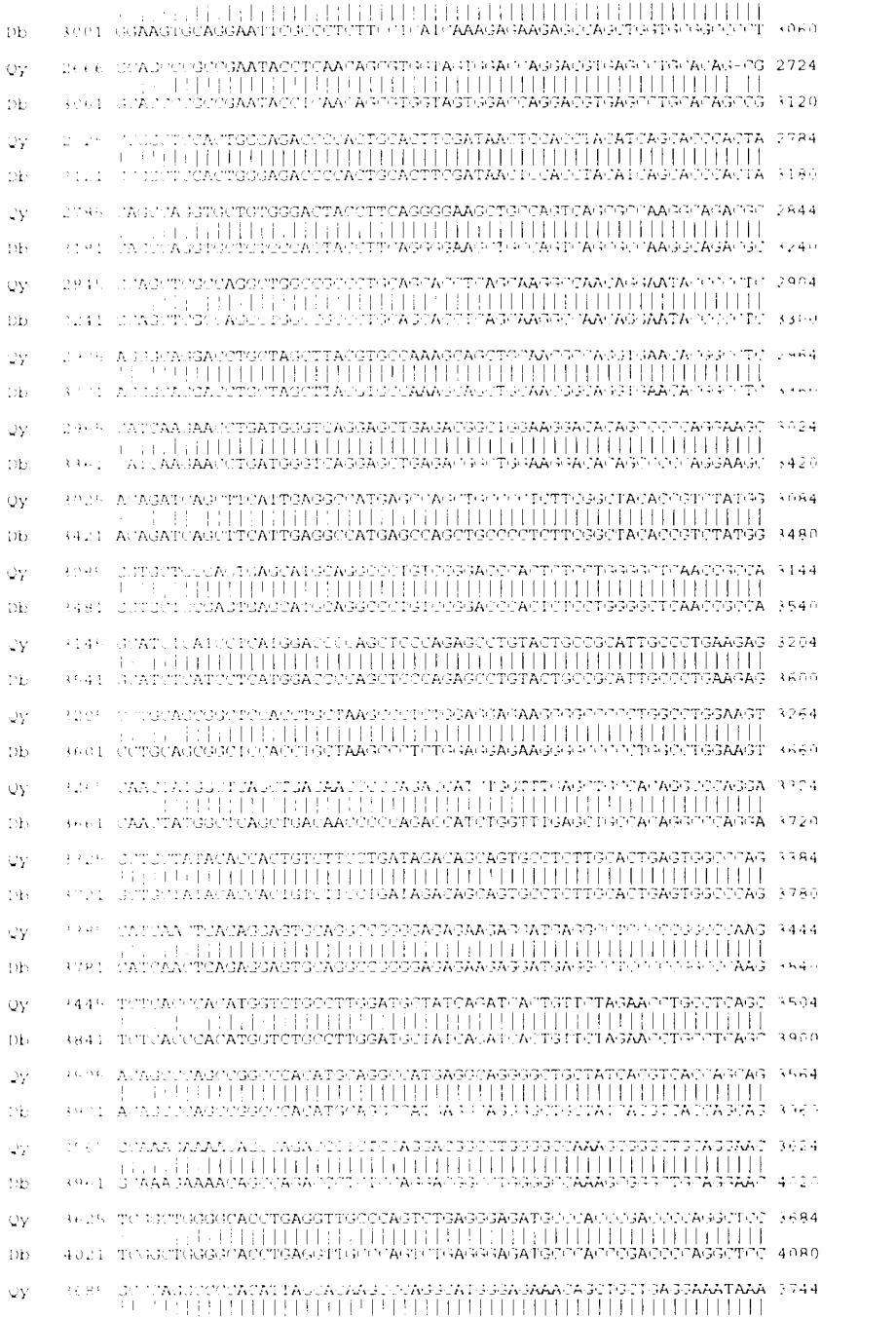
RESULT 2
AAS18821
ID AAS18821 standard; cDNA; 4174 BP.
XX
AC AAS18821;
XX
DT 19-DEC-2001 (first entry)
XX
DE DNA encoding human myosin-related protein, hMRP #1
XX
KW Human; oestrogen regulated unconventional myosin related protein; MRP;
KW auditory; antiatherosclerotic; osteopathic, neuroprotective; neutropic;
KW antigalactosemian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 561..1874
FT /*tag= a
FT /product= "hMRP"
FT /note= "Oestrogen-regulated unconventional myosin
FT related protein"
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-US08060.
XX
PR 17-MAR-2000; 2000US-188488P.
PR 09-MAR-2001; 2001US-0804156.
XX
PA (SCHE) SCHERING AG.
XX
PI Brosker AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-532747/58.
DR P-PSDB; AAU97400.
XX
PT Novel oestrogen-regulated unconventional myosin-related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS Claim 4; Fig 5; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating a wide variety of oestrogen related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the coding
CC sequence of human MRP variant #1 as described in the invention.
XX
SQ Sequence 4174 BP; 901 A; 1239 C; 1237 G; 697 T; 0 other;



Query Match 88.7%; Score 4352; DB 1; Length 4174;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 3778; Conservative 0; Mismatches 0; Indels 396; Gaps 4;







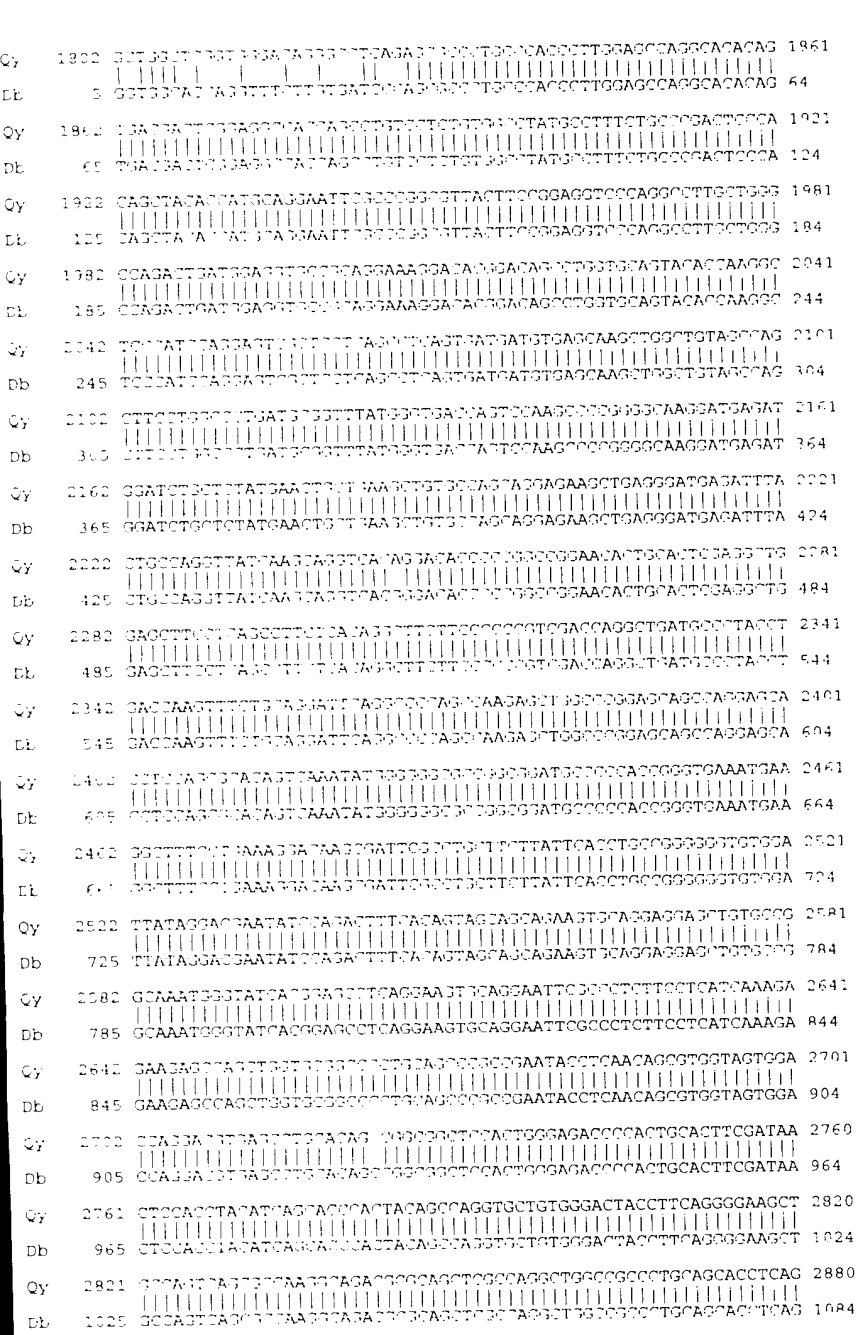
Db 4081 GCCCAGGCCCCACATTAGCACAAGCCCAGCATCCACAAAGAAA 4124

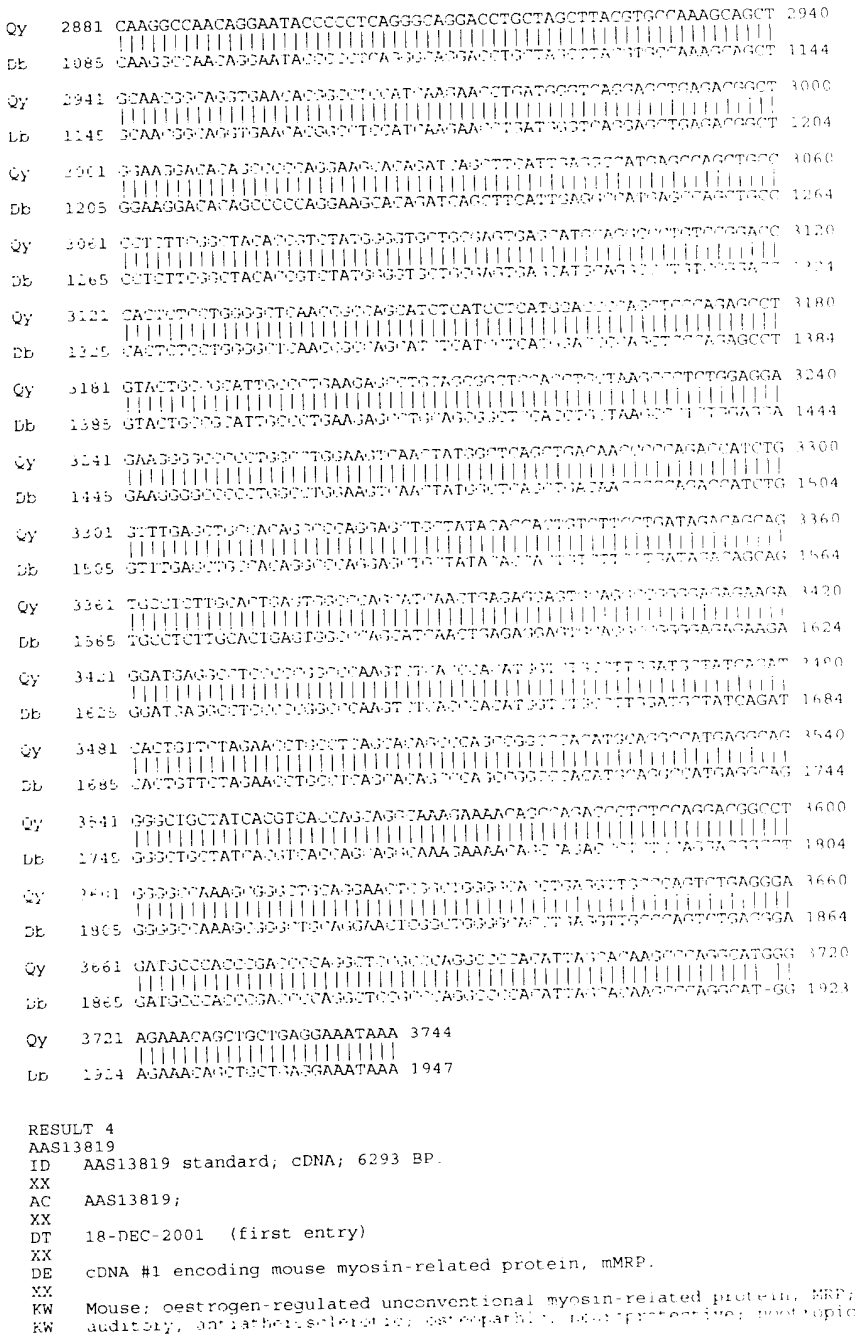
Qy 3745 ACTCCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3778

Db 4141 ACTCCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4174

RESULT 3
ABK35127
ID ABK35127 standard; cDNA; 1960 BP.
XX
AC ABK35127;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA encoding secreted protein #265.
XX
KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
KW tissue regeneration; wound healing; burn; haematopoiesis;
KW myeloid cell deficiency; lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN WO200177288-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO US10224.
XX
PR 06-APR-2000; 2000US 195582P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Bowes SH, Fesnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-179321/23.
XX
PT Five hundred and ninety two polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT -
XX
PS Claim 1; Page 202-203; 372pp; English.
XX
CC The invention relates to 592 polynucleotides which have been derived from
CC a variety of human tissue source and which encode novel secreted
CC proteins. The polynucleotides can be used as probes for the
CC identification and isolation of full length cDNA and genomic DNA. The
CC polynucleotides and proteins can also be used as nutritional supplements.
CC The proteins are useful in the treatment of various immune deficiencies
CC and disorders such as viral infections, bacterial infections, fungal
CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC and conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease, liver fibrosis, [illegible] ulcers. They are also
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment of
CC burns, incisions and ulcers. The proteins are also useful for regulating
CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.
XX
SQ Sequence 1960 BP; 433 A; 832 C; 567 G; 328 T; 0 other;

Query Match 50.1%; Score 1894.4; DB 24; Length 1960;
Best Local Similarity 99.1%; Pred No. 0;
Matches 1926; Conservative 0; Mismatches 16; Indels 2; Gaps 2;





RESULT 4
AAS13819
ID AAS13819 standard; cDNA; 6293 BP.
XX
AC AAS13819;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA #1 encoding mouse myosin-related protein, mMRP.
XX
KW Mouse; oestrogen-regulated unconventional myosin-related protein, MRP;
KW auditory, antiatherosclerotic; osteopathic, bronchoprotective; nootropic;

```
KW    antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW    osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW    Alzheimer's disease; Parkinson's disease; chromosome 11; immunogenic.
XX
OS    Mus sp.
XX
FH    Key             Location/Qualifiers
FT    CDS             125..4138
FT                    /*tag = a
FT                    /product= "mMRP"
FT                    /note= "oestrogen regulated unconventional myosin
FT                    related protein"
XX
PN    WO200168866-A2.
XX
PD    20-SEP-2001.
XX
PF    12-MAR-2001; 2001WO-US08060.
XX
PR    16-MAR-2000; 2000US-188488P.
PR    09-MAR-2001; 2001US-0803126.
XX
PA    (SCHD ) SCHERING AG.
XX
PI    Brooke AE, Deng GC, Fabbanyi CM;
XX
DR    WPI; 2001-632747/68.
DR    P-PSDB; AAU07399.
XX
PT    Novel oestrogen regulated unconventional myosin-related protein and
PT    polynucleotide for modulating effects of oestrogen in mammalian cells to
PT    treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS    Claim 4; Fig 2; 108pp; English.
XX
CC    The invention relates to an isolated oestrogen-regulated unconventional
CC    myosin related protein (MRP). Modulating MRP is useful for modulating the
CC    effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC    mammalian cell. When the level of expression or activity of MRP is
CC    increased, the development of atherosclerosis or osteoporosis is
CC    inhibited and when the level is decreased the development of breast
CC    cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC    useful for detecting the presence of oestrogen signaling in a mammalian
CC    cell, by detecting the expression or activity of MRP or the level of MRP
CC    mRNA in the cell. The presence of oestrogen signaling is used to
CC    determine the responsiveness of the cell to the oestrogen or the tissue-
CC    specific distribution of oestrogen signaling in a mammal. MRP is useful
CC    as a marker gene for the investigation of tissue-specific and oestrogen
CC    receptor specific ligands such as oestrogen receptor antagonists and
CC    agonists. Modulators, recombinant forms, derivatives, variants or
CC    fragments of MRP are useful for enhancing or inhibiting oestrogen
CC    signaling in cells and for treatment of breast cancer or other cancers
CC    and for preventing or treating wide variety of oestrogen-related
CC    diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC    disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC    preventing and/or treating hearing loss and deafness. MRP polypeptides
CC    are useful for generating monoclonal and polyclonal antibodies useful for
CC    identifying oestrogen responsive cells, particularly cells responsive to
CC    oestrogen through an oestrogen receptor alpha or beta, to identify
CC    alleles, interspecies homologues and polymorphic variants and in
CC    immunoassays to detect MRP. The present sequence represents the coding
CC    sequence of mouse MRP variant #1 as described in the invention.
XX
SQ    Sequence 6293 BP; 1433 A; 1917 C; 1733 G; 1210 T; 0 other;

  Query Match             39.4%;  Score 1490;  DB 22;  Length 6293;
  Best Local Similarity   67.1%;  Pred. No. 3.4e-284;
  Matches 2518;  Conservative    0;  Mismatches  810;  Indels  422;  Gaps   12;

Qy       4 CAGCAGCAGGCTGGGGCCTCGGAGGCTGCGTCCCAGGCCTCACCTCAGCCGTCACCTCC 63
           |||| |||| |||| |||| |||  |||||| |  |||||||| ||    |  ||
Db    2297 CAGTAGCAAGCTCAGACCTCTGGGGCCAGTCCCAGCCTCCACCTCAACCACTGCTCCC 2356
```

```
Qy      64 AAGGCCAAGAAGCCTCCTCCCCCAAGAAGCCAGAGAGTAACCTAGAGCCTTCCGGT 119
           ||| ||| |||||  ||   | ||||  |||||||| || ||  |  |   |  |||
Db    2357 AAGGCCAAGAAGCCTCCTCCCCCAAGAAGCCAGAGAGTAACCTAGAGCCTTCCGGT 2416

Qy     124 GGCTGCCTGAG---GAGAGCCTCCGAGGAGGCTGAAGACAGCCCTATCAGCCAAGAG 178
           |   ||  ||||    ||| ||| | ||||| ||||||  |  |||  | |  ||||||
Db    2417 GTTGGCTTGAGAGAGGACACCCAGAGGAAGCTGAAAGCAAGCCTCAGCCAAGAG 2472

Qy     181 TTCCAGCAGAAACCGAACTATTCCAGAGGATGCCCCACCAGTCAACTCAGCAG 234
           ||||| |||||||| |||||||||| |||||||||   | ||||| |||| ||||
Db    2473 TCTAAGAAAAAGCCGAATAATTCAAGAGATCCCCAAGAGGAGATCAGCAGCAG 2526

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGAAGCCCATGGAGAGGAGGAGAG 300
           |||| || || |||||||| || |||||||| | |||   | ||||  | ||||||||
Db    2527 GTGAAACCTCCAGCCAAGGTTCAGATCCCCAAGAGGAGATCAGCAGCAGCAGAG 2586

Qy     301 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAGTAGCCCTCCTCCCT 360
           ||||| |||    | |||  |                                | ||  | |||
Db    2587 GAGGATGAGAGCCCCGAGTTG---  --  -                    TCCCTCTCCCT 2634

Qy     361 CCTCCTCCCCCCATCCTGAAGAAGCCATTGAAGCAAGTGTGCCTCAAAGTCCCAAAGAG 420
           ||  |||||||  | ||||||||||  |||||  | |  || ||||||:  ||| ||
Db    2635 CCCCTCCCCCCGTTTGTGAAGAAGCCCGTGAAGCAAGCAGC---AAAGCCCTAAAGAA 2691

Qy     421 GCTGAGGCTGAGCCAGCCAAGGAG --  -  -                    A  440
           | ||||||  |||||  ||| |||||                             |
Db    2692 GATGAGGCAGAGCCCGCCCAGGAGGAAGTACCGACCCAGGCCGAGATCCCCCGTGCAG 2751

Qy     447 AGCCGCCAAGGGCCATGCCCAAGAGCCAGCCCAAGGAGGCCGCCCCGCCAGCCA 500
           |||  | |    ||   |  ||| ||| ||  |  |  |  |  |  |  ||
Db    2752 AGCTCCAACTCTGCACCTCAGCACCCCAAACCCAGCAGGTTACCCCAGCAGCAG 2807

Qy     507 GACTCCAAGCCCAAGCGGCCACAACCCAGCAGGGAAATTGCCAACACTATCGCATTTAT 566
           |||||   ||    || |||||||||||||||||  | ||||||||| ||||||||
Db    2813 AACTCCGCAGTTCCACGCCCCCAACCCAGCAGCAGAAATCCAAGTCATCAGCATTTAT 2872

Qy     567 CAGAGCCCCCGGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 626
           |||||||| || || || ||| |||||||||| || ||  |||||  |||| ||||
Db    2873 CAGAGCCGTCCAGGGCCTGTGGCTGTGCCGTACAACCCAGCAAGCCCATCAAAACTTCC 2932

Qy     627 CTGAGAAAAATCGACCCCAAGGAGCAGGCTCCCCCCAAGCCCAGCCAGCCCAGCCAGCC 686
           | ||| |||||  |||| |||| | | | | | | | | | | | | | | | | | |
Db    2933 CAGAAGAAAAATGACCCTAAGGATGAGGCTTTGGCTAAGCTTAAGAATAAATGGCGTCCA 2992

Qy     687 TTGCCCCCCCGATCCCCCCCAGCCCAAAAGCCCCCCCCCCCCCCCCCCCCTTGCCCTCCT 746
           | | |||  |||     || || | || ||| ||| | | | | | | | | | | |||
Db    2993 TTGCCCCTATCGA----CATCGCCTAACCAAGGAAGAGCTCCTCCACTCGTCTATCCCC 3048

Qy     747 CGACCCCAAGCCCCCCCCTACACCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 806
           ||||| ||||| | |  |  |||| ||| ||    |||||| ||||||||||||||||
Db    3049 CGACTTAAGTCCAACTCCTCCGAGCTCCCCCTACCCCCCCCCCCCCCCCCCCCCCCCC 3108

Qy     807 CTCCTCGAGCCTGTTTGGC--------  -AGAAGCTGCCTATCCCCACCACCCCCCCCT 857
           |||| |||| ||||||||           | || | |||| |  |  |  | | | ||
Db    3109 CTCCGGGACTTGTTTGCCCCATGTACTCCCAAACCCCACCTACCCCCCCCCCCCCCCT 3168

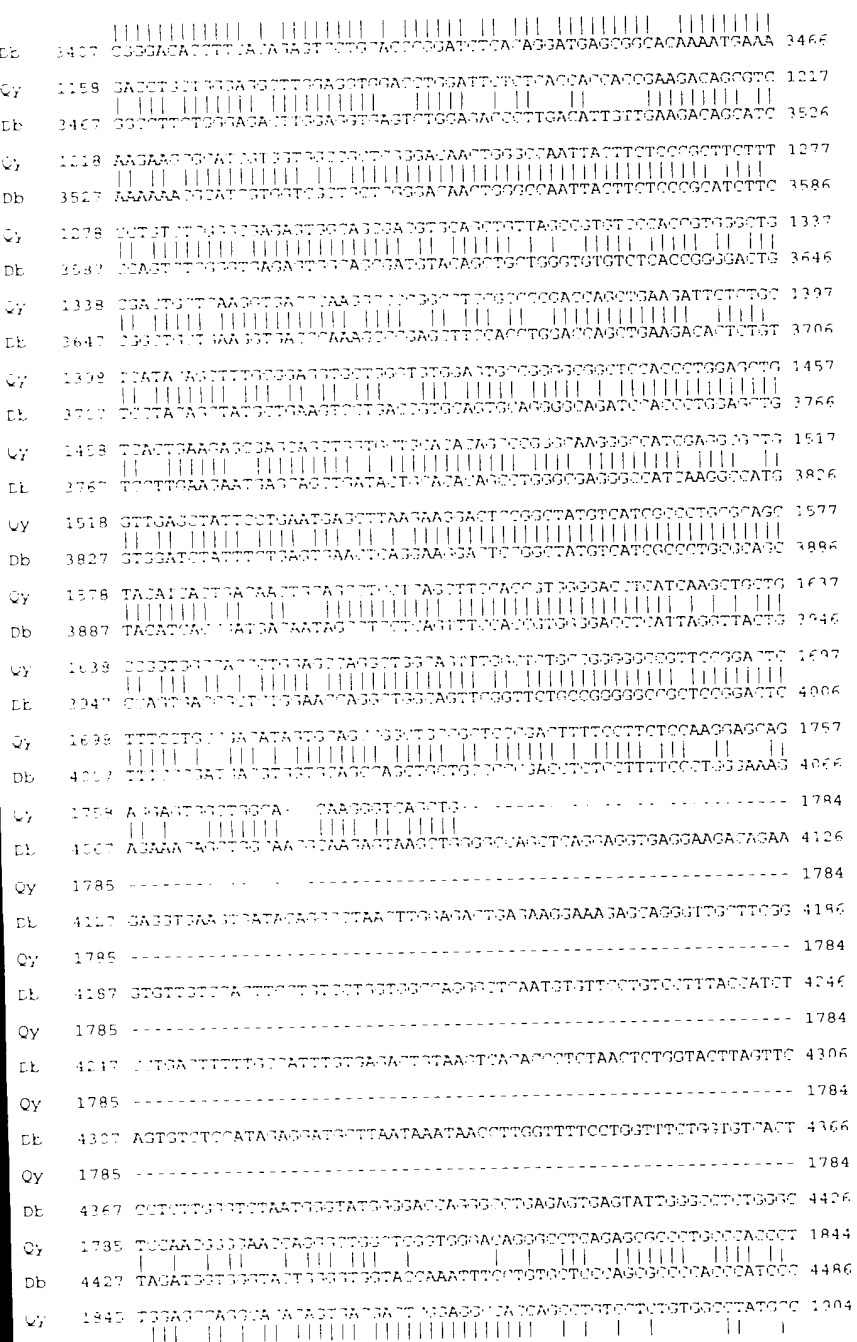
Qy     858 CCACCAGCCCCCACCACTGCCTCCTCCCCGAGGACCCAGGGACCCCCCCACCACCCCCCC 917
           |||||||| |  |||| ||||||| | | ||| ||   ||||| ||||| |||  | |
Db    3169 CCACCAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 3228

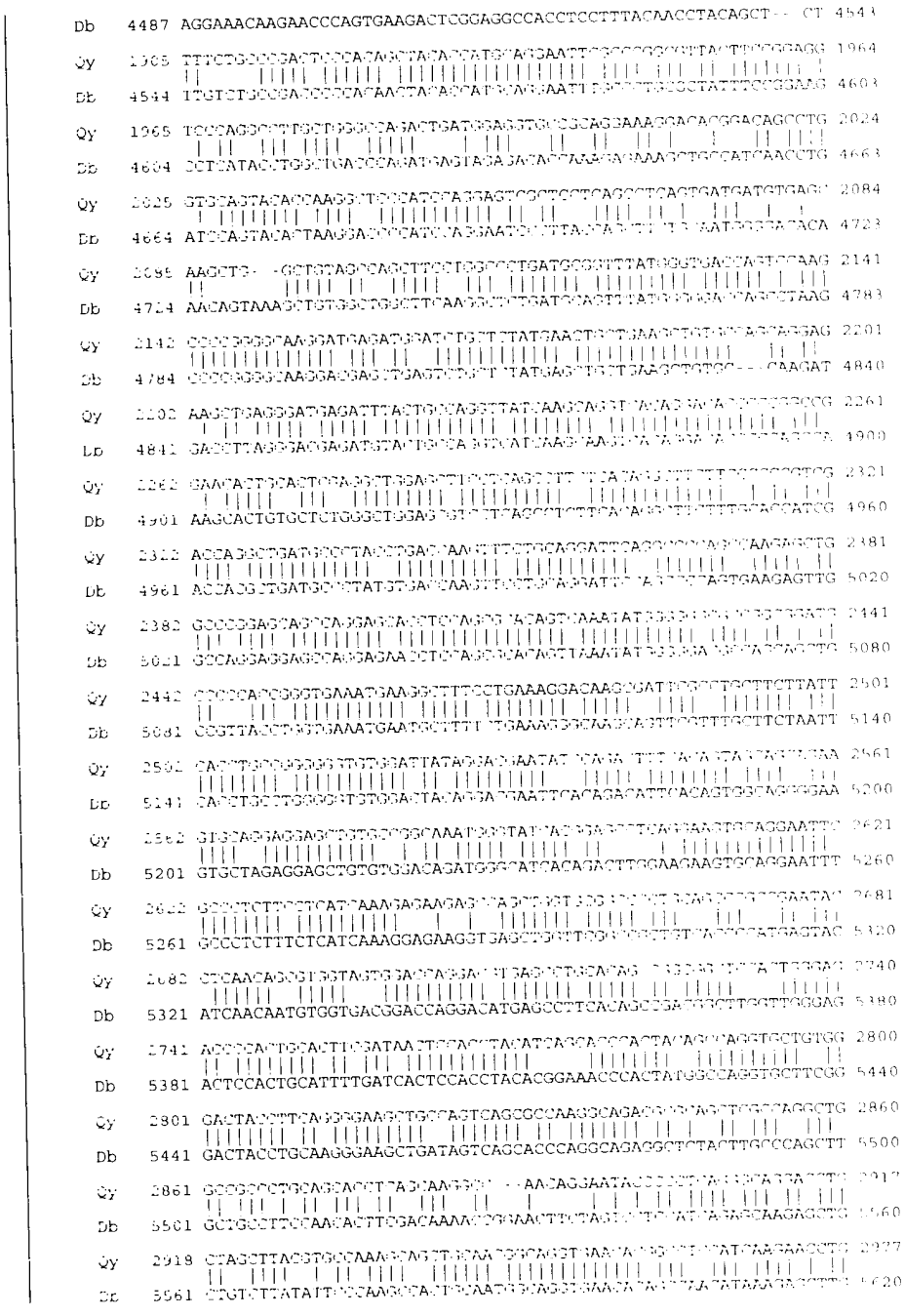
Qy     918 TGCCTGACCACAGCCCCCCGGAGGACCACAGGAGCCCCCCACCCAGCCCACCCCCCCCC 977
           |  |||||| |||||  |||||||  ||||||  |||| | |  |||||||||||||||
Db    3229 GCCCTGACAGAGCCCCATGGAGGACCAAGAACCATCCCCACCAAACCCCCCCCCCCCCC 3288

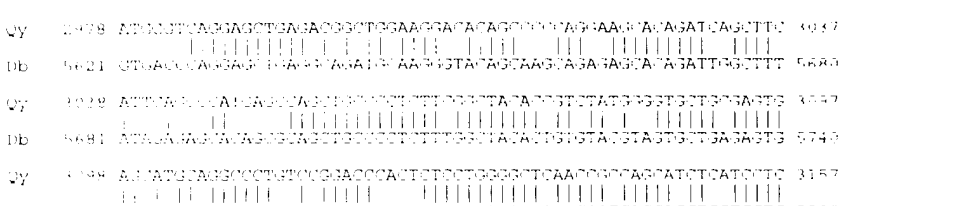
Qy     978 AGCCGCCGCCTTCCCCCACCACCGCCACGCCCTGGAAGTCGTCCCTACCCAAAGAGTGTTC 1037
           ||  |||||||||||||||  ||    | ||||||||||||||| || ||| ||||||||
Db    3289 AGTGTGTGCTTCTCCTATGCCAATGCACCCTGGAAGTCGTCCTACCCCAAAGAGTGTTC 3348

Qy     1038 TACCCCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCCCCCCCCCAGCAGATCCCTA 1097
            ||||| ||||||||||||||| |||| |||| ||| || ||||  |  ||||||||||
Db    3349 TACCCCCCGGGAGAACTTCAGCCATCCCTACTACCTGAGGCCCCCCCCCAGCAGATCCCTA 3408

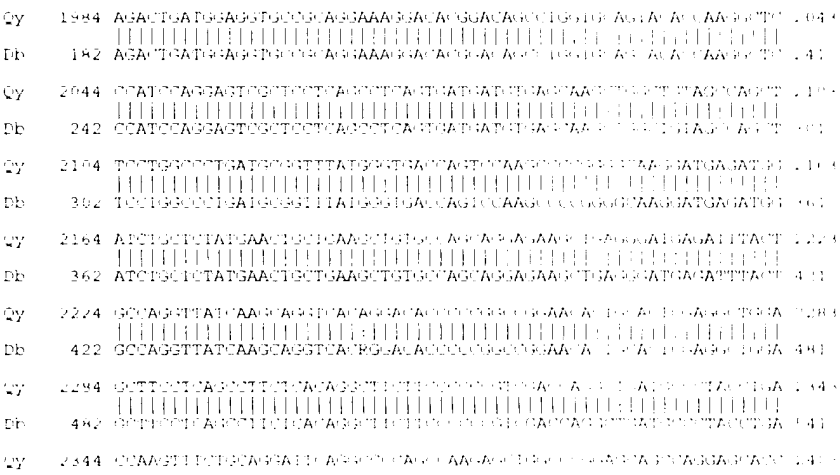
Qy     1098 CGGGACACCTTCCCCGAGTCCTGTACCCGAITTCCCAAAAGCCAAA 1142
            |||||||||||||||||||||||||||||||||||||||||||||
Db    3409 CGGGACACCTTCCCCGAGTCCTGTACCCGATTTCCCAAAAGCCAAA 3453
```

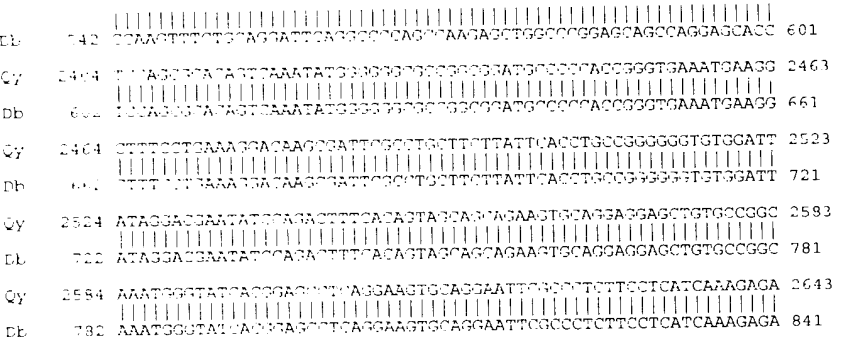






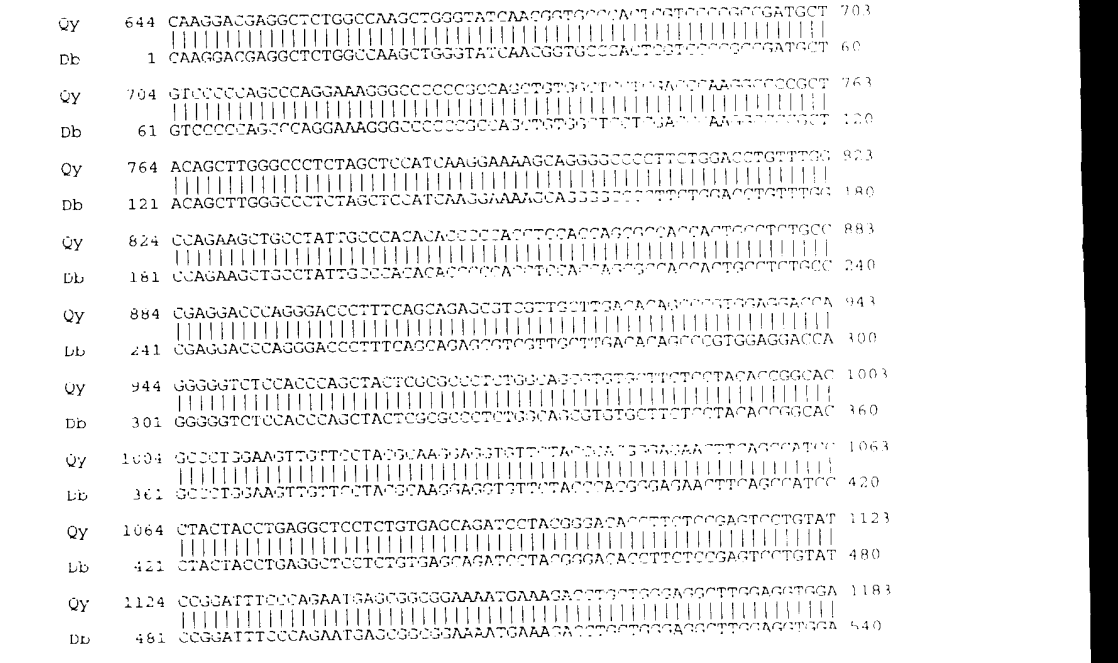
RESULT 5
AAS83274
ID AAS83274 standard; cDNA; 1346 BP.
XX
AC AAS83274;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 233.
XX
KW Immunomodulatory, human immunodeficiency virus, HIV, anaemia, angina,
KW rheumatoid arthritis; antiarteriosclerotic, cardiant, vascular,
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder, infection,
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX
OS Homo sapiens.
XX
PN WO200155329-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01347.
XX
PR 31-JAN-2000; 2000US-0179061.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451931/48.
DR P-PSDB; AAU20565.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
XX
PS Claim 1; SEQ ID No 233; 792pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example- immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio /
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS83043-AAS83486 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1346 BP; 294 A; 410 C; 407 G; 229 T; 6 other;

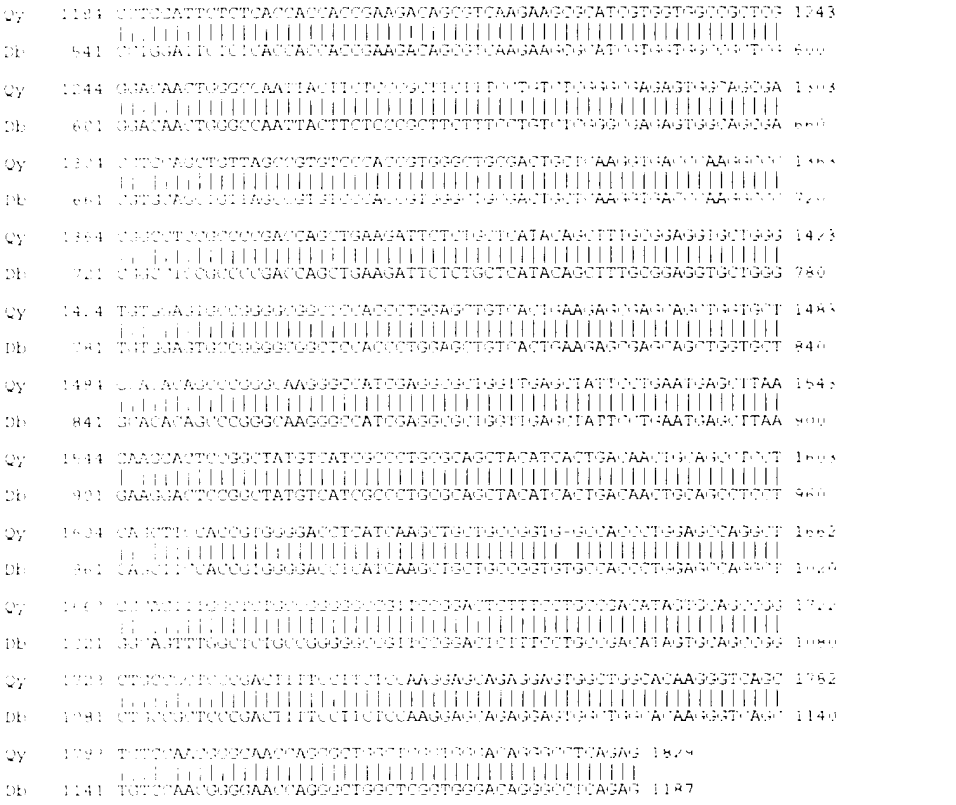




RESULT 6
AAZ56716
ID AAZ56716 standard; cDNA; 1476 BP
XX
AC AAZ56716;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN 19 encoding cDNA.
XX
KW Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder; ss
XX
OS Homo sapiens.
XX
PN WO9901471-A2.
XX
PD 07-JAN-1999
XX
PF 28-MAY-1998; 99WO-US011374
XX
PR 29-MAY-1998; 98US-0087260.
PR 02-JUL-1998; 98US-0091674.
PR 02-OCT-1998; 98US-0102954.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
PI Au-Young J,
XX
DR WPI; 2000-072605/06
DR P-PSDB; AAY57895.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders -
XX
PS Claim 3, Page 189, 229pp; English.
XX
CC AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN-1 to HTMPN 79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or
CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.
XX
SQ Sequence 1476 BP; 277 A; 479 C; 435 G; 285 T; 0 other;

Query Match 31.1%; Score 1175; DB 21; Length 1476;
Best Local Similarity 99.3%; Pred. No. 2.1e-222;
Matches 1186; Conservative 0; Mismatches 0; Indels 1; Gaps 1;



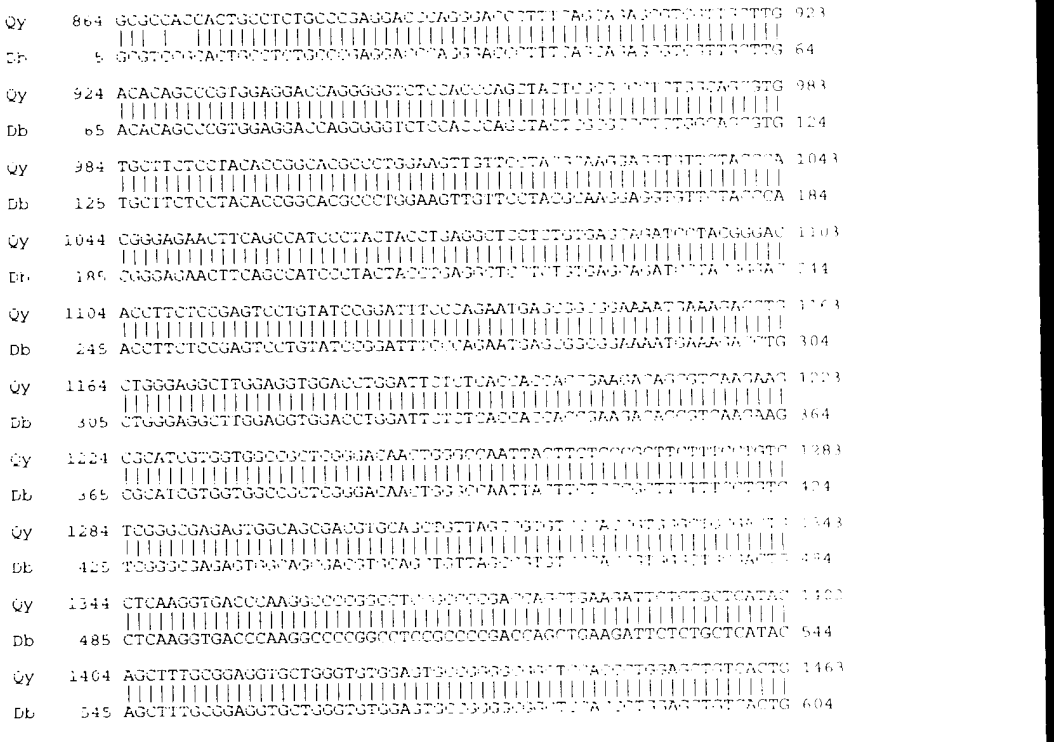


RESULT 7
ABK41544
ID ABK41544 standard; cDNA; 1268 BP.
XX
AC ABK41544;
XX
DT 05-JUN-2001 (first entry)
XX
DE DNA encoding novel central nervous system protein #124.
XX
KW [illegible] cell, system, CNS, inflammatory disease, rheumatoid arthritis,
KW hyperproliferative disorder, neoplasm, cardiovascular disorder,
KW cardiac arrest, cerebrovascular disorder, ischaemia, angiogenesis;
KW nervous system disorder, Alzheimer's disease, AIDS, ocular disorder;
KW acquired immunodeficiency virus, dysplasia, gastrointestinal disorder,
KW adenocarcinoma; reproductive system disorder, testicular feminisation,
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder, renal disorder, kidney failure; blood disorder;
KW myocardial infarction, wound healing, cell proliferation, skin aging,
KW food additive, food preservative; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200155318 A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.

XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217467.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224618.
PR 14-AUG-2000; 2000US-0224619.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225798.
PR 14-AUG-2000; 2000US-0225799.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229809.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250201.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251473.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251883.
PR 08-DEC-2000; 2000US-0251890.
PR 11-DEC-2000; 2000US-0254227.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR P-PSDB; AAU97214.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 1; SEQ ID No 134; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 25.4%; Score 961.2; DB 22; Length 1268;
Best Local Similarity 99.7%; Pred. No. 4.2e-180;
Matches 963; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



```
Qy    1464 AAGAGCAGCAGCTGGTGCTGCACACAGCCCCGGGCAAGAGCACATCGAGCGCTGGTTGAG 1523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     965 AAGAGCAGCAGCTGGTGCTGCACACAGCCCCGGGCAAGAGCACATCGAGCGCTGGTTGAG 924

Qy    1524 CTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCTGGGCAGCTACATC 1583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     665 CTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCTGGGCAGCTACATC 724

Qy    1584 AAGCAACTGGACGCCGACGAGCTCCAGAGCTTCGTGAACGAGCTGCATCAAGGTGGCTGGGAG 1643
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     725 AAGCAACTGGACGCCGACGAGCTCCAGAGCTTCGTGAACGAGCTGCATCAAGGTGGCTGGGAG 784

Qy    1644 ACCATGGAGCCAGGCTGGCAGTTTGGCTCTGGCAGCAGCGGCTCCATCGACACTCTTTCCTT 1703
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     785 ACCATGGAGCCAGGCTGGCAGTTTGGCTCTGGCAGCAGCGGCTCCATCGACACTCTTTCCTT 844

Qy    1704 CTGGACATAGTGCAGCGGGGCATGCCGCTCCGCATTTCCTTCCTCAAGGAGCAGAGGAGCT 1763
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     845 CTGGACATAGTGCAGCGGGGCATGCCGCTCCGCATTTCCTTCCTCAAGGAGCAGAGGAGCT 904

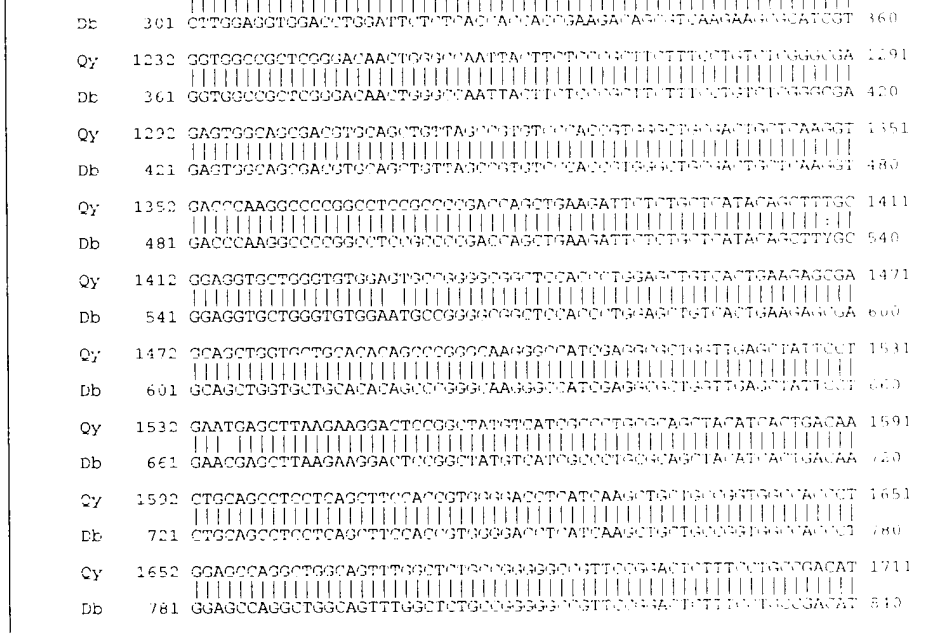
Qy    1764 GGCTGCCACAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGAGAGGGGT 1823
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     905 GGCTGCCACAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGAGAGGGGT 964

Qy    1824 TCAGAG 1829
           ||||||
Db     965 TCAGAG 970
```

RESULT 8
AAS27113
ID AAS27113 standard; cDNA; 1247 BP.
XX
AC AAS27113;
XX
DT 08-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 145
XX
KW Neuroprotective; cytostatic; dermal [illegible], immunosuppressive, [illegible],
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality, Down syndrome, ischaemia, renal disorder,
KW cardiovascular, respiratory, wound healing, endocrine, Addison's disease,
KW reproductive system, gastrointestinal, liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200149733-A1.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179261.
PR 04-FEB-2000; 2000US-0180028.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190078.
PR 18-APR-2000; 2000US-0198133.
PR 19-MAY-2000; 2000US-0205211.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215159.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217483.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225779.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234723.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250180.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251040.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0252719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR P-PSDB; AAU17193
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID No 145, 880pp, English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

Query Match 25.2%; Score 954.4; DB 22; Length 1247;
Best Local Similarity 99.7%; Pred No. 9.3e-179;
Matches 955; Conservative 1; Mismatches 2; Indels 0; Gaps 0;



```
Qy    1111 AGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGGAGTGGCTGGCA 1171
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGGAGTGGCTGGCA 900

Qy    1172 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAG 1231
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAAGGGTCAGCTGTCCAACGGGGAACCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAG 960


RESULT 4
ABK43853
ID   ABK43853 standard; cDNA; 1247 BP.
XX
AC   ABK43853;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding novel central nervous system protein #433.
XX
KW   Central nervous system; CNS; autoimmune disease, rheumatoid arthritis,
KW   hyperproliferative disorder, neoplasm, cardiovascular disorder,
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   [illegible] disease; Alzheimer's disease; AIDS; ocular disorder,
KW   impaired immunodefficiency; virus; dysphagia; gastrointestinal disorder;
KW   adrenocortical; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
DC   [illegible]19 AD.
XX
PD   [illegible] AUG 2001.
XX
PF   17-JAN-2001; 2001WO-US001332.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   [illegible]-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205521.
PR   07-JUN-2000; 2000US-0209467.
PR   08-JUN-2000; 2000US-0210896.
PR   [illegible]-JUN-2000; 2000US-0211335.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   [illegible]-JUL-2000; 2000US-0220063.
PR   26-JUL-2000; 2000US-0220984.
PR   14-AUG-2000; 2000US-0224618.
PR   14-AUG-2000; 2000US-0224619.
PR   14-AUG-2000; 2000US-0224615.
PR   14-AUG-2000; 2000US-0224714.
PR   14-AUG-2000; 2000US-0224768.
PR   14-AUG-2000; 2000US-0224847.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225767.
PR   14-AUG-2000; 2000US-0225768.
PR   14-AUG-2000; 2000US-0225769.
PR   18-AUG-2000; 2000US-0226279.
PR   [illegible]-AUG-2000; 2000US-0226681.
PR   [illegible]-AUG-2000; 2000US-0226858.
PR   [illegible]-AUG-2000; 2000US-0227192.
PR   [illegible]-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   [illegible]-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   [illegible]-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   [illegible]-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
```

PR 17-NOV 2000; 2000US-0249211.
PR 17-NOV 2000; 2000US-0249212.
PR 17-NOV 2000; 2000US-0249213.
PR 17-NOV 2000; 2000US-0249214.
PR 17-NOV 2000; 2000US-0249215.
PR 17-NOV 2000; 2000US-0249216.
PR 17-NOV 2000; 2000US-0249217.
PR 17-NOV 2000; 2000US-0249218.
PR 17-NOV 2000; 2000US-0249244.
PR 17-NOV 2000; 2000US-0249245.
PR 17-NOV 2000; 2000US-0249264.
PR 17-NOV 2000; 2000US-0249265.
PR 17-NOV 2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV 2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250231.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251038.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251884.
PR 08-DEC-2000; 2000US-0251930.
PR 11-DEC-2000; 2000US-0254041.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR P-PSDB; AAM98523.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 1; SEQ ID NO 443; 937pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyllitrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 26.2%; Score 954.4; DB 23; Length 1247;
Best Local Similarity 99.7%; Pred. No. 9.3e-179;
Matches 955; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 872 ACTGCCTCTGCCCAGGACCCAGGGACCTTTCAGCAGAGCGTCGTTGCTTGACACAGCC 931
Db 1 ACTGCCTCTGCCCAGGACCCAGGGACCTTTCAGCAGAGCGTCGTTGCTTGACACAGCC 60

Qy 932 CGTGGAGGACCAGGGGGTCTCCACCCAGCTACCTGCCCCCTCTGGCAGCCCCTGCTTCTC 991
Db 61 CGTGGAGGACCAGGGGGTCTCCACCCAGCTACCTGCCCCCTCTGGCAGCCCCTGCTTCTC 120

Qy 992 CTACACCGGCAGGCCCTGGAAGTTGTTCCTACCCAAGAAGAGTGCTTCTACCCATCCCAA 1051
Db 121 CTACACCGGCAGGCCCTGGAAGTTGTTCCTACCCAAGAAGAGTGCTTCTACCCATCCCAA 180

Qy 1052 CTTCAGCCATCCCTACTACCTGAGGCTCCTCCTGGAGCAGATCCTACCGGACACCTTCTC 1111
Db 181 CTTCAGCCATCCCTACTACCTGAGGCTCCTCCTGGAGCAGATCCTACCGGACACCTTCTC 240

Qy 1112 CGAGTCCTGTATCCGGATTCCCAGAATGAGGGGCGAAAATGAAAGACCTGCTGGGAGG 1171
Db 241 CGAGTCCTGTATCCGGATTCCCAGAATGAGGGGCGAAAATGAAAGACCTGCTGGGAGG 300

Qy 1172 CTTGGAGGTGGACCTGGATTCTCTCAGCAGCAGCGAAGACAGCCTCAAGAAGCCCATCGT 1231
Db 301 CTTGGAGGTGGACCTGGATTCTCTCAGCAGCAGCGAAGACAGCCTCAAGAAGCCCATCGT 360

Qy 1232 GGTGGCCGCTGGGACAACTGGGCCAATTACTTCTTCCTTTTCCTGTCTCGGGCGA 1291
Db 361 GGTGGCCGCTGGGACAACTGGGCCAATTACTTCTTCCTTTTCCTGTCTCGGGCGA 420

Qy 1292 GAGTGGCAGCGACGTGCAGCTGTTAGCCTTCTTCCAGCAGCGGCTGTTACTGCTCAAGGT 1351
Db 421 GAGTGGCAGCGACGTGCAGCTGTTAGCCTTCTTCCAGCAGCGGCTGTTACTGCTCAAGGT 180

Qy 1352 GACCCAAGGCCCCGGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 1411
Db 481 GACCCAAGGCCCCGGCCTCCGCCCCGACCAGCTGAAGATTCTCTGCTCATACAGCTTTGC 540

Qy 1412 GGAGGTGCTGGGTGTGGAATGCGGGGGCGGCTCCACCCTGGAGCTGTCACTGAAGAGCGA 1471
Db 541 GGAGGTGCTGGGTGTGGAATGCGGGGGCGGCTCCACCCTGGAGCTGTCACTGAAGAGCGA 600

Qy 1472 GCAGCTGGTGCTGCACACAGCCCCGGCAAGGCCATCGAGGCGCTGGTTGAGCTATTCCT 1531
Db 601 GCAGCTGGTGCTGCACACAGCCCCGGCAAGGCCATCGAGGCGCTGGTTGAGCTATTCCT 660

Qy 1532 GAATGAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCCTGCCGCAGCTACATCACTGACAA 1591
Db 661 GAACGAGCTTAAGAAGGACTCCGGCTATGTCATCGGCCCTGCCGCAGCTACATCACTGACAA 720

Qy 1592 CTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGCCCGTGGCCACCCT 1651
Db 721 CTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTGCCCGTGGCCACCCT 780

Qy 1652 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGCGCCGTTCCGCACCTTTTCCTGCGACAT 1711
Db 781 GGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGCGCCGTTCCGCACCTTTTCCTGCGACAT 840

Qy 1712 AGTGCAGCGGGCTGCGGCTGCCGACTTTCCTCTCCAAGGAGCAGAGGAGTGGCTGGCA 1771
Db 841 AGTGCAGCGGGCTGCGGCTGCCGACTTTCCTCTCCAAGGAGCAGAGGAGTGGCTGGCA 900

Qy 1772 CAAGGGTCAGCTGTCCAAGGGGAACCAGGGCTGGCTGCCTGCGACAGGCCTCAGAG 1829
Db 901 CAAGGGTCAGCTGTCCAAGGGGAACCAGGGCTGGCTGCCTGCGACAGGCCTCAGAG 960

RESULT 10
AAS13820
ID AAS13820 standard; cDNA; 4375 BP.
XX
AC AAS13820;
XX
DT 18-DEC 2001 (first entry)
XX
DE cDNA #2 encoding mouse myosin-related protein, mMRP.
XX
KW Mouse; oestrogen-regulated unconventional myosin-related protein; MRP;

```
KW    auditory; antiatherosclerotic; osteopathic; neuroprotective; nootropic;
KW    antiprotozoan; cytostatic; oestrogen antagonist; atherosclerosis;
KW    osteoporosis; breast cancer; cardiovascular disease; deafness; ss;
KW    Alzheimer's disease; Parkinson's disease; Huntington's chorea.
XX
OS    Mus sp.
XX
FH    Key             Location/Qualifiers
FT    CDS             125..4138
FT                    /tag= a
FT                    /product= "mMRP"
FT                    /note= "Oestrogen-regulated unconventional myosin-
FT                          related protein"
XX
PN    WO200168866 A2.
XX
PD    20-SEP-2001.
XX
PF    12-MAR-2001; 2001WO-US08060.
XX
PR    10-MAR-2000; 2000US-188488P.
PR    29-MAR-2000; 2000US-0804126.
XX
PA    (MYRI-) MYRIAD GENETICS INC.
XX
PI    Brooks AR, Dong GG, Rubanyi GM;
XX
DR    WPI; 2001-622747/68.
DR    P-PSDB; AAU07499.
XX
PT    Novel oestrogen-regulated unconventional myosin-related protein and
PT    polynucleotide for modulating effects of oestrogen in mammalian cells to
PT    treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS    Claim 4; Fig 3; 108pp; English.
XX
CC    The invention relates to an isolated oestrogen-regulated unconventional
CC    myosin related protein (MRP). Modulating MRP is useful for modulating the
CC    effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC    mammalian cell. When the level of expression or activity of MRP is
CC    increased, the development of atherosclerosis or osteoporosis is
CC    inhibited and when the level is decreased the development of breast
CC    cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC    useful for detecting the presence of oestrogen signaling in a mammalian
CC    cell, by detecting the expression or activity of MRP or the level of MRP
CC    [mRNA or the] cell. The presence of oestrogen signaling is used to
CC    determine the responsiveness of the cell to the oestrogen or the tissue-
CC    specific distribution of oestrogen signaling in a mammal. MRP is useful
CC    as a marker gene for the investigation of tissue-specific and oestrogen
CC    receptor specific ligands such as oestrogen receptor antagonists and
CC    agonists. Modulators, recombinant forms, derivatives, variants or
CC    fragments of MRP are useful for enhancing or inhibiting oestrogen
CC    signaling in cells and for treatment of breast cancer or other cancers
CC    and for preventing or treating wide variety of oestrogen-related
CC    diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC    disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC    prevention and/or treating hearing loss and deafness. MRP polypeptides
CC    are useful for generating monoclonal and polyclonal antibodies useful for
CC    identifying oestrogen responsive cells, particularly cells responsive to
CC    oestrogen through an oestrogen receptor alpha or beta, to identify
CC    alleles, interspecies homologues and polymorphic variants and in
CC    immunoassays to detect MRP. The present sequence represents the coding
CC    sequence of mouse MRP variant #2 as described in the invention.
XX
SQ    Sequence 4375 BP; 966 A; 1368 C; 1216 G; 825 T; 0 other;

  Query Match             25.0%;  Score 946.4;  DB 22;  Length 4375;
  Best Local Similarity   75.9%;  Pred. No. 4.7e-177;
  Matches 1312;  Conservative    0;  Mismatches  416;  Indels   76;  Gaps    5;

Qy       4 CAGAAAACAAGCTCGAGGCCTCGGAGGCTGGCTCCAAGGCCTCAGCCTCAGTGGTCACCTCC 63
           || | || |||| | |||| | ||| | ||||||| |  |||||||| ||    |  ||
Db    2297 CAGGAACAAAGCTCAGACCTCTGGGGCCAGCTCCCAGCCTCCAACCTCAACCACTGCTCCC 2356
```

```
Qy      64 AAGCCCAAGAAGCCCCCCACAACCCCCGAGAACCAACAGCCCAGCCCTCGAGTCAGAGAGC 123
           ||| ||| |||||| ||  | ||||  |||||||| || |  ||| |  |   ||||
Db    2357 AAGTCTAAGAAGTCTTCTTCTCTCTCAGAGAACTAAGAAGTCTAAGAGTCTCAGTCTTTCCT 2416

Qy     124 GCCTGCCTGAG--GGAGAGCTCCCAGCAGCCCGAAGACAGCCTCTATCAGCCTAAGACTTA 181
           |  || ||||   ||| ||| | |||||  ||||||  || ||| | |  ||||||||
Db    2417 GTTGGCTTGAGAGAGGACAGCCCAGAGGAAGTGAAAGTAAGTCTAAGTCCCAAGAGT 2475

Qy     182 TTCCAGCAGAAACGGAACTATTTCCAGAGATGCCCCACCTAAGATCAATGAAGACG 240
           |||| |||||||| ||||||||||| ||||||||  . |||| |||| ||||
Db    2477 TTCCAACAGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCTGATCAAGATGAAGACG 2536

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGAAGCGCCAGAAGAAGAGAGAGAG 300
           |||| || || |||||||| || |||||||| | ||  | |  | |||| |||||
Db    2537 GTGAAACCTCCAGCCAAGGTTCAGATCCCCCAAGAGAAGATGAAGAAGAAGAAGAAG 2596

Qy     301 GAGGAGGAGGAGGAGGAGCAGGAGGAGCAAGAAGTGGAAACAAGAAGCAGCCGCCTCCCT 360
           ||||| |||   |  |||  |                    . ||  | ||
Db    2597 GAGGATGAGACCGCCGAGTTG----  -  -                TCCCTCTCTCCT 2629

Qy     361 CCTCCTCCCCCCATGTGGAAGAAGCCATTGAAGCAAAGTCTCCCCAAAGCTCCAAAAGAG 420
           || |||||||| | ||||||||||  ||||| | |  || || ||||  ||| ||
Db    2630 CCCCCTCCCCCCGGTTGTGAAGAAGCCGGTGAAGCCAAGCATCCAAAGCCTTAAAGAAA 2689

Qy     421 GCTGAGGCTGAGCCAGCCAAGGAG  --  --                          AC 466
           | ||||| |||| ||| ||||                                     ||
Db    2690 GATGAGGCAGAGCCCGCCAGAGAAGTACCGACCCAAGA--AAGAGCCCGCCTCCTCCAC 2747

Qy     467 AGCGGCCAAGGGCCATGGCCAAGGCCAGCCAAGCCAAGCCAAGCCAGTCCCAGTTA 524
           |:| | |    ||  |  ||| || ||  ||  |    |  |  |  | |
Db    2750 AGTTCCAACCTCCGCACTCAGCACTCCAAACCCAGCAGCACCCA  CACTCCAGCAGTTCC 2809

Qy     525 GACTCCAAGCCCAAGCCGCCCACAACCCAGCAGGGAAATCGCCAAGATCATCATCGAAGTAC 584
           |||||  ||    || || ||||||||||||||||    || |||||| || |||
Db    2810 AACTCCGCACCTCCACCGCCCGCAACCCAGCAGGGAAATCGAAACATCATCCTGAAGTCAC 2869

Qy     585 CAGAGCCCCCCCGGGCCCCTCTGCCTCTCCCCGTCGCAGCCAGCCAGCCAGCCTTTCCAAA 644
           |||||||| || || || |||||||||| || ||  |||  |  |  |  |  |||||
Db    2870 CAGAGCCCTCCAGCCCCTCTCCCTCTCCCTAGAAACCACCAGCCCAGCCAAAAATCTTT 2929

Qy     645 CTGAGGAAAATCGACCCCAAGGACGAGCCTCTCGCCAAGCTCCCAGCAACCGTCCCCAC 704
           | || |||||  |||||  |||||  |||||  |||| ||  |  |  ||  || | ||||
Db    2930 CAGAAGAAAAATGACCCTAAGGATGAGCCTTTCGCTAAGCTACCCGACAAAGTGCCTCCAC 2989

Qy     705 TCCCCCCCCCCCGACCCCCCCCCCCCAGCCCAAAAAGCCC--  --  ---CCCCCCCCCC 760
           || |||  |||     || || | || || ||||  |||||  |  |   |||||||||| |||
Db    2990 TTGCCCCTATCGA--CATCGCCTAACCAAAGGAAGACCCCCCAGCCTTAACTTCCCT 3046

Qy     761 CGACCCAAGGCCCCCCTACAGCCCCCCCCCCCCAGCCCCCCCCCCAAAAAAGCAGCCCCCC 820
           |||||  ||||| |  |  |  |||| ||| ||   ||||||  ||||||||||||| |||
Db    3047 CGACCTAAGCCTCCACCTCCCCCCACCCCCCCCAGCCCCCCCCCCAAAAACAGCGACCCC 3106

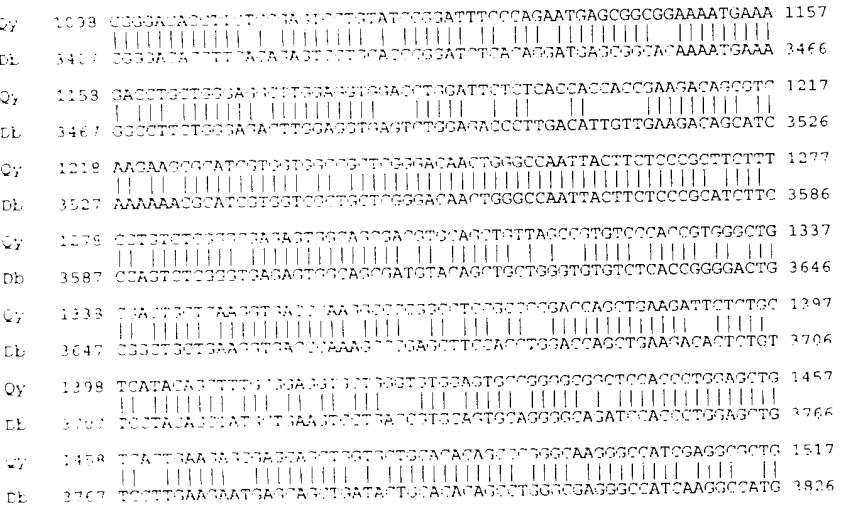
Qy     821 CTTCTCGACCTGTTTGGC          CACAACCCCCCCACCCC  CCCCCCCCCCCCCCC 877
           |||| |||| |||||||  |  ||  |  |||  |||  ||  ||||||| ||||
Db    3107 CTTCTGGACTTGTTTGGCCCACTGTAGTCCAAACCCACCTACCCCTCCCCCCCCTC 3165

Qy     878 CCACCCAGCCCCACCACTGCCCTCTGCCCGAGGACCCAGGGACCCCTTCCACCAGAGCGTCGT 937
           |||||||  |  |||| |||||| | | ||| ||   |||  |||  ||  |||| ||||
Db    3167 CCACCCACCACTCCCACCCCCCTCTGTCTCCCCGACCCCAACCACCTTCACCCAGACGTCTCAT 3226

Qy     938 TGCTTGACACAGCCCCTGGAGGACCAGGGGCTCTCCACCCAGCTACTCCCCCCCTCTCCCC 997
           |||||||  |||| |||||||| ||   |||||||  |||| ||  ||  |||| ||||
Db    3227 GCCTTGACACAGCCCCATCGAGGACCAAGAACCATCTCCCAAAAGCCTCCTTCCCCCTCCCA 3286

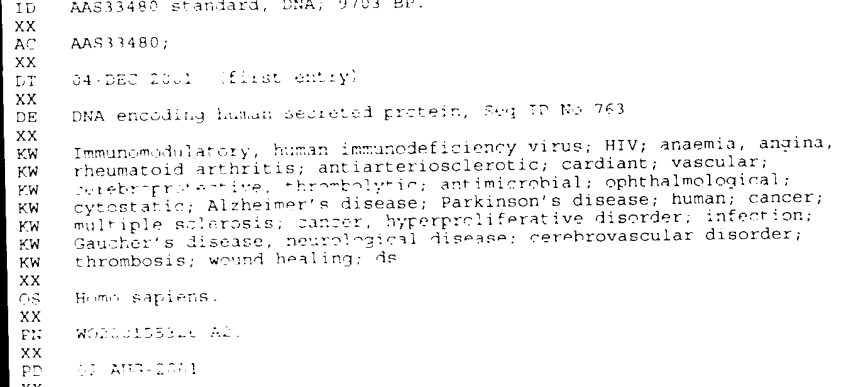
Qy     998 AGCCTGTCCTTCTCCCTACACCCCCACCCCCTCCCCCTGGAAGTTGTCCCTACTCCAAGTAGTTTTT 1057
           || |||||||||||||  ||   | ||||||||||||||| |||| ||||||||||||
Db    3287 AGTCTCTCCCCTTCCCTATCCCCAAGCCCACCCCCCAAAGCCCCCCCCCCCCCCCCCCC 3346

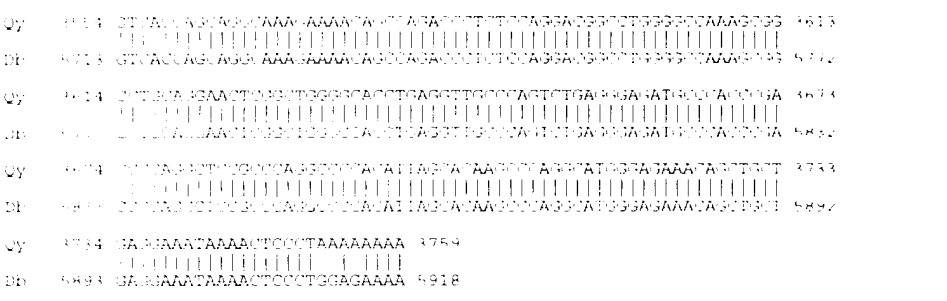
Qy     1058 TACCCACCGGGAGAACTTCAGCCATCCCTACTACCTGACGCTCTTCTTGACTAGTATTTA 1117
            |||| ||||||||||||| |||| |||| ||| ||  |||||||||||||||||||||
Db    3347 TACCCCCGGGAGAACTTCAGTCATCCATACTGCTTCACTTCTTCCTCTTCACTAGATTTTA 3406
```



RESULT 11
AAS33480
ID AAS33480 standard; DNA; 9703 BP.
XX
AC AAS33480;
XX
DT 04-DEC 2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 763
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ds
XX
OS Homo sapiens.
XX
PN WO200155211-A2.
XX
PD 02-AUG-2001
XX
PF 17-JAN-2001; 2001WO-US01347
XX
PP 31-JAN-2000; 2000US-0179065.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451931/48.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
XX
PS Disclosure; SEQ ID No 763; 753pp; English
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and of which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example, immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9703 BP; 1950 A; 2842 C; 3079 G; 1832 T; 0 other;

Query Match 11.7%; Score 442, DB 22; Length 9703;
Best Local Similarity 92.1%; Pred. No. 1.6e-77;
Matches 466; Conservative 0; Mismatches 40; Indels 0; Gaps 0;





RESULT 12
AAS27687
ID AAS27687 standard; DNA; 24757 BP
XX
AC AAS27687;
XX
DT 07-NOV-2001 (first entry)
XX
DE DNA encoding novel signal transduction pathway protein, Seq ID 1347
XX
KW Neurological, cytostatic, hematological, immunosuppressive, tumour,
KW antiinflammatory, anti-HIV, antibacterial, antiinflammatory, cancer,
KW immune system disorder; rheumatoid arthritis, inflammatory condition,
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular, respiratory, wound healing, endocrine, Addison's disease;
KW reproductive system; gastrointestinal, liver disorder, AIDS, ds;
KW acquired immune deficiency syndrome
XX
OS Homo sapiens
XX
PN WO200147733 A1.
XX
PD 05 AUG 2001.
XX
PF 17 JAN 2001; 2001WO-US01412
XX
PR 21 JAN 2000; 2000US-0179065
PR 24 FEB 2000; 2000US-0184664
PR 16-MAR-2000; 2000US-0189874
PR 17 MAR 2000; 2000US-0190076
PR 18 APR 2000; 2000US-0198123
PR 19 MAY 2000; 2000US-0205515
PR 07 JUN 2000; 2000US-0209467
PR 28 JUN 2000; 2000US-0214986
PR 30 JUN 2000; 2000US-0215155
PR 07 JUL 2000; 2000US-0216647
PR 07 JUL 2000; 2000US-0216880
PR 11-JUL-2000; 2000US-0217487
PR 11 JUL 2000; 2000US-0217496
PR 14-JUL-2000; 2000US-0218240
PR 26 JUL 2000; 2000US-0220983
PR 26 JUL 2000; 2000US-0220984
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519
PR 14-AUG-2000; 2000US-0225114
PR 14 AUG 2000; 2000US-0225214.
PR 14 AUG 2000; 2000US 0225268.
PR 14-AUG-2000; 2000US-0225267
PR 14 AUG 2000; 2000US-0225268
PR 14 AUG 2000; 2000US 0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30 AUG 2000; 2000US 0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08 SEP 2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14 SEP 2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14 SEP 2000; 2000US-0232399.
PR 14 SEP 2000; 2000US-0232400.
PR 14 SEP 2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14 SEP 2000; 2000US 0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21 SEP 2000; 2000US 0234274.
PR 25 SEP 2000; 2000US-0234997.
PR 25 SEP 2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27 SEP 2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US 0236327.
PR 29-SEP-2000; 2000US 0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US 0237040.
PR 13-OCT-2000; 2000US 0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08 NOV 2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US 0246528.

PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250191.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251388.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251888.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254337.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID NO 1347; 888pp; English
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AA826976 AA827850 represent novel signal transduction
CC pathway protein-encoding sequences and PCR primers of the invention
XX

SQ Sequence 24757 BP; 5183 A; 7437 C; 7328 G; 4809 T; 0 other;

Query Match 11.7%; Score 442; DB 22; Length 24757;
Best Local Similarity 95.1%; Pred. No. 2.1e-77;
Matches 466; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 3254 GGCCTGGAAGTCAACTATGGCTCAGCTGACAACCCCTAGATCATCTGTTTGAGCTGCCA 3313
Db 4289 GGCCGGGGCAGAGGCGAGGGGCTGGCTGGGGACCCTCTTCCTAATAGCCCTTTCTTCCGG 4230

Qy 3314 CAGGCCCAGGAGCTGCTATACACCACTGTCTTCCTGATAGACAACAGTGCCTCTTTGACT 3373
Db 4229 CAGGCCCAGGAGCTGCTATACACCACTGTCTTCCTGATAGACAACAGTGCCTCTTTGACT 4170

Qy 3374 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGCCCGGGAGAGAAGAGGATGAGCCCTCC 3433
Db 4169 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGCCCGGGAGAGAAGAGGATGAGCCCTCC 4110

Qy 3434 CCCGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 3493
Db 4109 CCCGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 4050

Qy 3494 CCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCCATGAAAAAGAGGGTGCTATCAC 3553
Db 4049 CCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCCATGAAAAAGAGGGTGCTATCAC 3990

Qy 3554 GTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTTCAAGACAAAAAGGGG 3613
Db 3989 GTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTTCAAGACAAAAAGGG 3930

Qy 3614 GCTGCAGGAACTCGGCTGGGGCACCTGACATTGGCCACTTGAAGGACATGCCCACCCGA 3673
Db 3929 GCTGCAGGAACTCGGCTGGGGCACCTGACATTGGCCACTTGAAGGACATGCCCACCCGA 3870

Qy 3674 CCCCAGGCTCGGCCCAGGCCCACATTAGACAAGCCCAGTCATCTGAAAAAGAGGTGGT 3733
Db 3869 CCCCAGGCTCGGCCCAGGCCCACATTAGACAAGCCCAGTCATCTGAAAAAGAGGTGGT 3810

Qy 3734 GAGGAAATAAAACTCCCTAAAAAAAA 3759
Db 3809 GAGGAAATAAAACTCCCTGGAGAAAA 3784

RESULT 13
AAS33481
ID AAS33481 standard, DNA, 24757 BP.
XX
AC AAS33481;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID NO 764.
XX
KW Immunomodulatory; human immunodeficiency virus, HIV; anaemia; angina;
KW rheumatoid arthritis, antiarteriosclerotic, cardiant; vascular;
KW cerebroprotective; thrombolytic, antimicrobial; ophthalmological;
KW cytostatic, Alzheimer's disease, Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer, hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease, cerebrovascular disorder;
KW thrombosis; wound healing; ds.
XX
OS Homo sapiens.
XX
PN WO200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN 2001; 2001WO-US01347.
XX
PR 31-JAN 2000; 2000US-0179065
XX
PA (HUMA-) HUMAN GENOME SCI INC
XX

```
PI   Fosser JA,  Parash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Disclosure; SEQ ID No 764; 758pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example, immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot Marie Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC   angina and thrombosis), infections caused by bacteria, viruses and
CC   fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC   agonists, antagonists and antibodies can also be used to promote wound
CC   healing, maintain organs before transplantation, and support cell culture
CC   of primary tissues. AAE28043 AAE28486 represent human secreted protein
CC   coding sequences, PCR primers, and related sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 24757 BP; 4809 A; 7328 C; 7437 G; 5183 T; 0 other;

  Query Match             11.7%;  Score 442;  DB 22;  Length 24757;
  Best Local Similarity   92.1%;  Pred. No. 2.1e-77;
  Matches  466;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy    4254 GCTCTTGCAAGTCAACTATGGCTCAGCTGACAACCCCCAGACCACTCTGGTTTGACCTGCCA 4313
           ||  ||        | | ||  |||    | |||       | | ||  |  |||||||||
Db   20469 GGCTGGCCGCAGAGGGAGGGGGTCTGGGGACCTTCTTTAACTAGGCTTTCTTGGG 20525

Qy    4314 CATGCCCAGGAGCCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTTGCACTT 4373
           ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20526 CATGCCCAGGAGCCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTTGCACTT 20585

Qy    4374 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAACAGGATGAAGCCTCC 4433
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20586 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAACAGGATGAAGCCTCC 20645

Qy    4434 CCAGGCCCAAGTCTCAGCCACACTGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 4493
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20646 CCAGGCCCAAGTCTCAGCCACACTGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 20705

Qy    4494 CCTTCCTCAGACACAGCCACGTGGCCACATGTAGGGTATGAGGTAGGGGCTGTTATCAC 4553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20706 CCTTCCTCAGACACAGCCACGTGGCCACATGTAGGGCATGAGGTAGGGGCTGTTATCAC 20765

Qy    4554 GTTAGCAAAAAAAAAAAAAAAAGCAAGAGCCTCTCCAGGAGGGCTGGGGCCAAAGGG 4613
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20766 GTTAGCAGAGGCAAAGAAAACAGCCAGACCCTCTCCAGGAGGGCTGGGGCCAAAGGG 20825

Qy    4614 GCTGCAGGAACTCGGCTGGGGCACCTGACGTTGCCCAGTCTCAAAGACATGCCCACCCGA 4673
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20826 GCTGCAGGAACTCGGCTGGGGCACCTGACGTTGCCCAGTCTCAAAGACATGCCCACCCGA 20885

Qy    4674 CCCCAGGCTCCGCCCAGGCCCACATTAGCACAAGCCCAGGCATGGAGAAACAGCTCCT 4733
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20886 CCCCAGGCTCCGCCCAGGCCCACATTAGCACAAGCCCAGGCATGGAGAAACAGCTCCT 20945

Qy    4734 GAGGAAATAAAACTCCCTAAAAAAAA 4759
           ||||||||||||||||||| ||||
Db   20946 GAGGAAATAAAACTCCCTTGAGAAAA 20974


RESULT 14
AAD07791
ID   AAD07791 standard; cDNA; 1044 BP.
XX
AC   AAD07791;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein encoding gene #  cDNA clone HECNO45, SEQ ID NO:..
XX
KW   Human, secreted protein, proliferative disorder, cancer, tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder, AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation, allergy, neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder, schizophrenia; asthma;
KW   skin disorder, psoriasis, sepsis, diabetes, atherosclerosis;
KW   cardiovascular disorder, angina, pain disorder, kidney disorder;
KW   gastrointestinal disorder, pregnancy related disorder;
KW   endocrine disorder, infection, wound healing, pulmonary;
KW   cell culture, chemotaxis, food additive, gene therapy;
KW   binding partner identification, etc.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             616..667
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT                   /note= "CDS does not include start codon."
FT                   /partial
XX
PN   WO200132687-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US29364.
XX
PR   29-OCT-1999;  99US-0162238.
PR   30-JUN-2000;  2000US-0610134.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Duan RD,  Fikes JD,  Florence M,  Fisher R;
XX
DR   WPI; 2001-308740/32.
DR   P-PSDB; AAE03399.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein, is
PT   used in preventing, treating or ameliorating a medical condition
XX
PS   Claim 1; Page 362; 498pp; English.
XX
CC   AAD07770-AAD07794 represent cDNAs corresponding to human secreted
CC   protein genes, and AAE03378 AAE03407 represent the proteins they encode.
CC   AAE03408-AAE03415 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   10 genes, based on the tissues in which they are most highly expressed,
```

CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1044 BP; 217 A; 350 C; 326 G; 146 T; 5 other;

Query Match 8.5%; Score 323; DB 22; Length 1044;
Best Local Similarity 98.4%; Pred. No. 2.8e-54;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 [illegible] 60
Db 721 [illegible] 779

Qy 61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120
Db 780 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 839

Qy 121 GGTGGCTGCTTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
Db 840 GGTGGCTGCTTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 899

Qy 181 TTCAAGAAAACGGAACTATTTCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
Db 900 TTCAGCAGAAACGGAACTATTTCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 959

Qy 241 ATGAAGCCCCCCCCCAAGGTCCACATCCCCCAGGGGAAGCGCAGGAGGAGGAGGAGGAG 300
Db 960 ATGAAGCCCCCCGCCAAGGTCCACATCCCCCAGGGGAAGCGCAGGAGGAGGAGGAGGAG 1019

Qy 301 GAGGAGGAGGAGGAGGAGCGGAGG 325
Db 1020 GAGGAGGAGGAGGAGGAGCGCGGAGG 1044

RESULT 15
ABN95022/c
ID ABN95022 standard; DNA; 311 BP.
XX
AC ABN95022;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #1520 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds, hepatocellular carcinoma, hepatotropic;
KW metastatic liver tumour; cytostatic, expression profile; disease state,
KW disease progression, drug toxicity, drug efficacy, drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1, SEQ ID NO 1520; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503 ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 311 BP; 44 A; 85 C; 103 G; 79 T; 0 other;

Query Match 7.7%; Score 292; DB 24; Length 311;
Best Local Similarity 98.1%; Pred. No. 2.7e-48;
Matches 306; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 3447 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 3506
Db 311 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 252

Qy 3507 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGCTGCTATCAGTCACCAGCAGGC 3566
Db 251 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGCTGCTATCAGTCACCAGCAGGC 192

Qy 3567 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGGAGGTGCAGGAACTC 3626
Db 191 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGGAGGTGCAGGAACTC 132

Qy 3627 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGAGATGCCCACCCGACCCCACG-TCCGC 3685
Db 131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGAGATGCCCACCCGACCCCACG-TCCGC 73

Qy 3687 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGAGAAAACAGCTGCTGAGGAAATAAAAC 3746
Db 72 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGAGAAAACAGCTGCTGAGGAAATAAAAC 13

Qy 3747 TCCCTAAAAAAA 3758
Db 12 TCCCTGGAGAAA 1

Search completed: January 9, 2003, 07:23:29
Job time : 890.674 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

SW nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:xx:47 ; Search time 4512.17 Seconds
(without alignments)
13567.518 Million cell updates/sec

Title: US-09-803-126-7
Perfect score: 3780
Sequence: 1 cggcagcgcgcagggtcgggt ... aaaaaaaaaaaaaaaaaaaa 3780

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	874.8	23.1	969	14	BM804816	BM804816 AGENCOURT
2	771.2	20.4	920	12	BG680906	BG680906 602628729
3	754	19.9	1004	14	BM804831	BM804831 AGENCOURT
4	735.8	19.5	1070	13	BI766913	BI766913 603053164
5	715.4	18.9	836	13	BG912704	BG912704 602907089
6	699.2	18.5	1015	14	BM923763	BM923763 AGENCOURT
7	687.6	18.2	908	13	BI766411	BI766411 603052657
c 8	683.2	18.1	689	14	BM977966	BM977966 UI-CF-EC1
9	681.8	18.0	728	14	BM720779	BM720779 UI-E-EO1-
10	606.4	16.0	661	13	BI769659	BI769659 603055555
11	603	16.0	603	14	BM791149	BM791149 K_EST0071
12	595.8	15.8	626	10	AW753302	AW753302 RC3-CT025
c 13	580	15.3	623	13	BG943621	BG943621 [illegible]
14	573.4	15.2	676	10	AV683102	AV683102 AV683102
c 15	564	14.9	567	9	AI825877	AI825877 [illegible]
16	559	14.8	877	12	BG426026	BG426026 602492704
c 17	545.2	14.4	637	13	BI046691	BI046691 MR3-FN020
c 18	543.6	14.4	591	12	BF817068	BF817068 MR2-CI012
c 19	533.2	14.1	538	14	BM996595	BM996595 UI-H-DT0
c 20	507.4	13.4	594	10	AW854251	AW854251 RC3-CT025
21	496.6	13.1	923	14	BQ956737	BQ956737 AGENCOURT
22	484.4	12.8	722	12	BI772088	BI772088 603059167
23	482.2	12.8	530	13	BI007177	BI007177 QV3-BT007
24	481.6	12.7	881	13	BG968181	BG968181 602835662
25	467.6	12.4	799	13	BI772324	BI772324 603056028
26	467.4	12.4	669	12	BF334334	BF334334 RC3-CT025
c 27	467	12.4	478	12	BF805323	BF805323 QV1-CI017
28	465.4	12.3	567	12	BG877603	BG877603 PM2-HT052
c 29	463	12.2	471	10	AW361379	AW361379 RC3-CT025
c 30	462.4	12.2	481	12	BF116088	BF116088 7n77d04.x
31	460.2	12.2	485	14	BQ374676	BQ374676 RC2-TN007
c 32	457.2	12.1	463	12	BF434203	BF434203 7p71b01.x
c 33	444.4	11.8	473	13	BM150946	BM150946 TCBAP1D12
c 34	444	11.7	452	10	AW299454	AW299454 ve60e11.x
35	439.8	11.6	494	13	BM152909	BM152909 TCBAP1Q12
36	434.4	11.5	444	14	BM702060	BM702060 UI-E-CQ1
37	433.2	11.5	448	13	BM666647	BM666647 UI-E-CQ1-
c 38	430.2	11.4	446	10	AW136317	AW136317 UI-H-BI1-
c 39	424.2	11.2	440	13	BM150833	BM150833 TCBAP1D12
c 40	418	11.1	423	12	BF434683	BF434683 7p03a07.x
41	417.8	11.1	526	10	AW753342	AW753342 RC3-CT025
c 42	417	11.0	715	14	BQ574071	BQ574071 UI-H-BZ0
43	416.8	11.0	504	9	AA448391	AA448391 zw68b09.r
44	413.8	10.9	942	12	BF577925	BF577925 602091945
45	413.6	10.9	474	14	BM792920	BM792920 K-EST0073

ALIGNMENTS

RESULT 1
BM804816
LOCUS BM804816 969 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT_6494843 NIH_MGC_175 Homo sapiens cDNA clone IMAGE:6025517 5', mRNA sequence.
ACCESSION BM804816
VERSION BM804816.1 GI:19121639
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 969)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I M A G E Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12716 row: i column: 06
High quality sequence stop: 626.
FEATURES Location/Qualifiers
source 1..969

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5725517"
/clone_lib="NIH MGC_125"
/lab_host="DH10B"
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site_2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yr. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

BASE COUNT 208 a 333 c 267 g 158 t 3 others
ORIGIN

Query Match 23.1%; Score 874.8; DB 14; Length 969;
Best Local Similarity 97.6%; Pred. No. 1.8e-109;
Matches 940; Conservative 0; Mismatches 18; Indels 5; Gaps 5;

Qy 1161 GACACCTTCTTCGAGTCTGTATCCGCATTTCCAGAATGAGCTGAAATTAAAATA 1220
Db 781 GACACCTTCTTCGAGTCTGTATCCGCATTTCCAGAATGAGCTGAAATTAAAATA 840

Qy 1161 CTGCTGG-GAGGCTTGGAGGTGGACCTGGATTCTCCACCACCAAGACACATCTCCAA 1219
Db 841 CTGCTGGGNAGGCTTGGAGGTGGACCTGGAATCTCCACCACCAAGACACATCTCCAG 900

Qy 1220 GAAGCCATTGTGTCCTTG.TGCAAA.AAGCAAGTTTTTGCTTTT 1278
Db 901 AAACCCCATCGTGGGTGCCGCTTCGGGACCACTA.TTCTCCGCTTCTTT 958

Qy 1279 CTG 1281
Db 959 CTG 961

RESULT 2
BG680906
LOCUS BG680906 920 bp mRNA linear EST 01-MAY-2001
DEFINITION 602628729F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4758625, mRNA sequence.
ACCESSION BG680906
VERSION BG680906.1 GI:13912816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 920)
AUTHORS NIH MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rmail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10613 row: i column: 08
High quality sequence stop: 889.
FEATURES Location/Qualifiers
source 1..920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4758625"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: 5'-CAGGAAACAGCTATGACCATG-3'. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 211 a 284 c 264 g 161 t
ORIGIN

Query Match 20.4%; Score 771.2; DB 12; Length 920;
Best Local Similarity 95.5%; Pred. No. 1.9e-95;
Matches 858; Conservative 0; Mismatches 35; Indels 7; Gaps 6;

Qy 2298 TTCTCACAGGCTTCTTCCCCCCGTGGACCAGGTGATGCCTACTCACTAAGTTTTTGC 2357
Db 13 TTTTACTGTCAACTTCCCCCCCTGCACCAGGCTGATGCCTACTCACAAGTTTTTGC 72

Qy 2358 AGGATTCAGGCCCCAGCCAAGAGCTGGCCGGAGCAGCCAGGAGCACCTCAGCCATAT 2417
Db 73 AGGATTCAGGCCCCAGCCAAGAGCTGGCCGGAGCAGCCAGGAGCACCTCAGCCAGAG 132

Qy 2418 TCAAATATGGGGGGGGGGGGGGGATGGCCCCATGGGGGAAATGAACTTCCCTAAAAG 2477

```
Db     133 TCAAAATATGGGGGGCA[illegible]ATGCCCTACCGGGGTGAAATGAAGGCTTTCCTGAAAG 192

Qy    2476 GAGAAG[illegible]GCTTCTTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATA 2535
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 GACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATA 252

Qy    2536 TCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTATCA 2595
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 TCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTATCA 312

Qy    2596 CGGAGCCTCAGAAAGTGCAGAATTCCCCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGG 2655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 CGGAGCCTCAGAAAGTGCAGGAATTCCCCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGG 372

Qy    2656 TGGAGGCTGAGAAGTGCAGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCC 2715
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 TGGAGGCTGAGAAGTGCAGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCC 432

Qy    2716 TGATAGG[illegible]GGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCA 2774
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     433 TGCATAGCCCGGGGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCA 492

Qy    2775 GCACCCACTACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAGCTGCCAGTCAGCGCCA 2834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     493 GCACCCACTACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAGCTGCCAGTCAGCGCCA 552

Qy    2835 AGGTAGACTGTGCTAGTGCCAGGCTGGCCCTGCAGCACCTCAGCAAGGCCAACAGGA 2894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     553 AGGCAGACGGGCAGCTGCCAGGCTGGCCCTGCAGCACCTCAGCAAGGCCAACAGGA 612

Qy    2895 ATACCGGCTACAGGTACGACTTCTAGTTTACGTGCCAAAGCAGCTGCAACGGCAGGTCA 2954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     613 ATACCGGCTACAGGTACGACTTCTAGTTTACGTGCCAAAGCAGCTGCAACGGCAGGTCA 672

Qy    2955 ACACGGGCTGCATCAAGAACCTGATGGCTCAGGAGCTGAGAACGGCTGGAAGGACACAGCC 3014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 ACACGTGCTGCATCAAGAACCTGATGGCTCAGGAGCTGAGAACGGCTGGAAGGACACAGCC 732

Qy    3015 CCCAGGAAGCACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGGCCCTCTTCGGCTAC 3074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     733 CCCAGGAAGCACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGGCCCTCTTCGGCTAC 792

Qy    3075 ACCGTCATATGGGTTGTTGCGAAGTGAGCATGCAGGCCGTGCCGGACCCACTCTCCTGG 3134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     793 ACCGTCATATGGGTTGTTGCGAAGTGAGCATGCAGGCCGTGCCGGACCCACTCTCCTGG 852

Qy    3135 GGGTCAACCCCTAGCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCG 3189
           ||  | ||    ||||||| ||||||||||    |||||||  |||||||| ||
Db     853 GGGCTCAA[illegible]GCATCTCAT-CTCATGGACCC--AGTCCCAGAAGCTGTACTGGCG 907
```

```
RESULT 2
BM804831
LOCUS       BM804831                1004 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6474845 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725541
            5', mRNA sequence.
ACCESSION   BM804831
VERSION     BM804831.1  GI:19121654
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation

            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13716  Row: j  Column: 16
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1..1004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5725541"
                     /clone_lib="NIH_MGC_125"
                     /lab_host="DH10B"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source (pool
                     of three ovaries, from females ranging in age from 38 to
                     43 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1.5-5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber [Invitrogen]. Reference Strausberg
                     tracking code 036."
BASE COUNT       231 a    347 c    269 g    157 t
ORIGIN
```

Query Match 19.9%; Score 754; DB 14; Length 1004;
Best Local Similarity 96.6%; Pred. No. 4e-93;
Matches 813; Conservative 0; Mismatches 25; Indels 4; Gaps 4;

```
Qy     322 GAGGAGCAAGAAGTGGAAACAAGAGCAGCCCGTCCCTCCTCCTCCCCCATCGTGAAG 381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGGAGCAAGAAGTGGAAACAAGAGCAGTCCCGTCCCTCCTCCTCCCCCATCGTGAAG 60

Qy     382 AAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAG 441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAG 120

Qy     442 GAGACAGCGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGACTGTGGTGGGC 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAGACAGCGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGACTGTGGTGGGC 180

Qy     502 AGTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGGCAAATTGGCAACATCATCCGC 560
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGCTCAGACTCCAAGCCCAAGCGGCCACAACCCAGCAGGCAAATTGGCAACATCATCCGC 240

Qy     561 ATGTACCAGAGCCGCCGGGCCCCGTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAA 620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATGTACCAGAGCCGCCGGGCCCCGTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAA 300

Qy     621 GCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGGTGGGTATCAACGGT 680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTTTCCTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGGTGGGTATCAACGGT 360

Qy     681 GCCCACTCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCAGGCAGTGTG 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCCCACTCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCAGGCAGTGTG 420

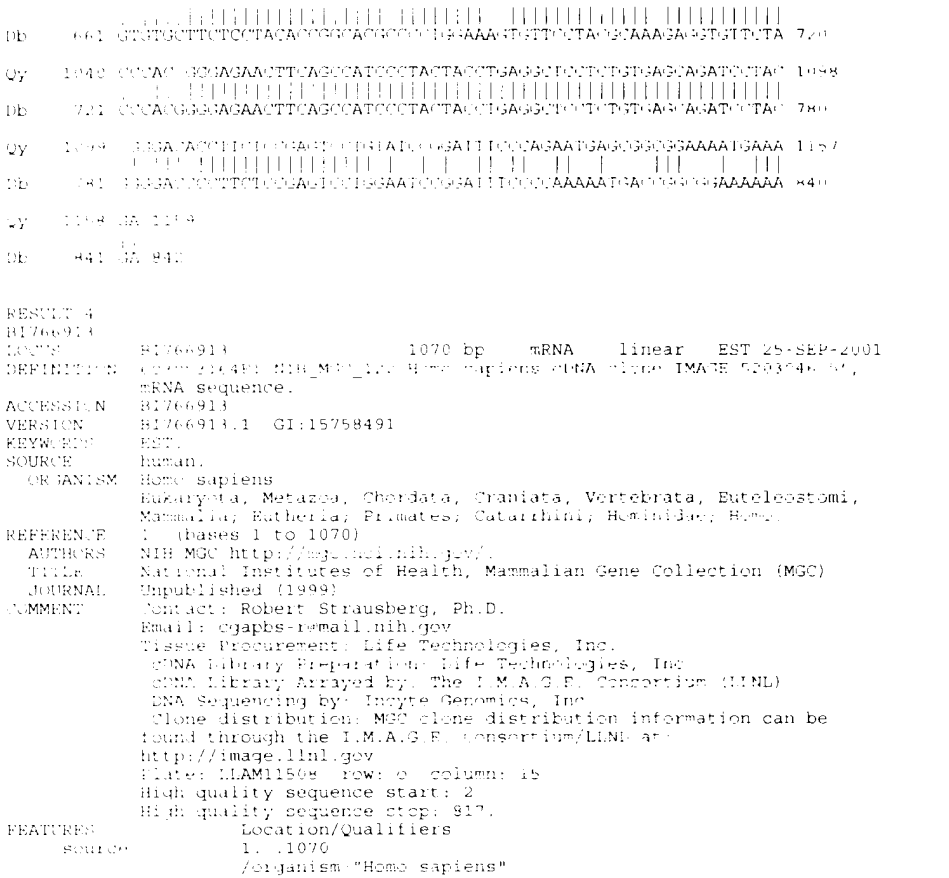
Qy     741 GCTCCTCGACCCAAGGCCCCGCTACAGCTTGGGCCCTCTAGCTCCATCAAGGAAAAGCAG 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GCTCCTCGACCCAAGGCCCCGCTACAGCTTGGGCCCTCTAGCTCCATCAAGGAAAAGCAG 480

Qy     801 GGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACACACCCCCACCTCCA 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGGCCCCTTCTGGACCTGTTTGGCCAGAAGCTGCCTATTGCCCACACACCCCCACCTCCA 540

Qy     861 CCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGC 920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCAGCGCCACCACTGCCTCTGCCCGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGTTGC 600

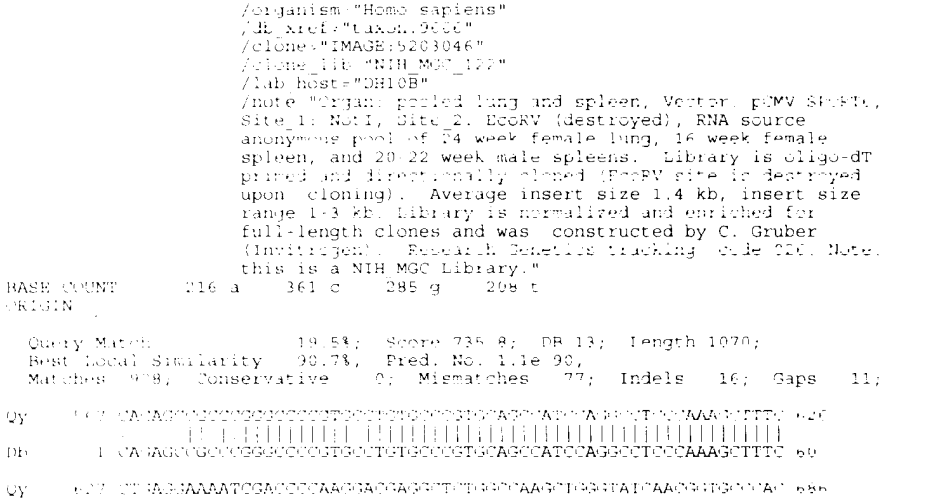
Qy     921 TTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGCCCCTTGTGCAGC 980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGCCCCTTGTGCAGC 660

Qy     981 GTGTGCTTCTCCTACACCGGCACG CCCTGGAAGTTGTTCCTACCGCAAGGAGCGTGTTCTA 1039
```



RESULT 4
BI766913
LOCUS BI766913 1070 bp mRNA linear EST 25-SEP-2001
DEFINITION [illegible] NIH_MGC_126 Homo sapiens cDNA clone IMAGE 5203046 5', mRNA sequence.
ACCESSION BI766913
VERSION BI766913.1 GI:15758491
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
REFERENCE 1 (bases 1 to 1070)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM1150g row: o column: 15
High quality sequence start: 2
High quality sequence stop: 917.
FEATURES Location/Qualifiers
source 1..1070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5203046"
/clone_lib="NIH_MGC_127"
/lab_host="DH10B"
/note "Organ: pooled lung and spleen; Vector: pCMV SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 220. Note: this is a NIH MGC Library."
BASE COUNT 216 a 361 c 285 g 208 t
ORIGIN

Query Match 19.5%; Score 735.8; DB 13; Length 1070;
Best Local Similarity 90.7%; Pred. No. 1.1e-90;
Matches 978; Conservative 0; Mismatches 77; Indels 16; Gaps 11;





RESULT 5
BG912704
LOCUS BG912704 836 bp mRNA [illegible] EST [illegible] JUN-[illegible]
DEFINITION [illegible] NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4[illegible], mRNA sequence.
ACCESSION BG912704
VERSION BG912704.1 GI:14293180
KEYWORDS EST.
SOURCE human.

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10875  row: h  column: 23
            High quality sequence stop: 827.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4938934"
                     /clone_lib="NCI_CGAP_Brn67"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      185 a    247 c    249 g    155 t
ORIGIN

  Query Match             18.9%;  Score 715.4;  DB 13;  Length 836;
  Best Local Similarity   91.7%;  Pred. No. 7.2e-88;
  Matches  828;  Conservative    0;  Mismatches    6;  Indels   69;  Gaps    4;

Qy   1974 GGCCACCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCACAGCTACACCAT 1933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      1 GGCCAUCAGCCTGTCCTCTGTGGCCTATGCCTTTCTGCCCGACTCCCACAGCTATACCAT 60

Qy   1934 GCAGGAATTCGCCCCGCCGTTACTTCCGGAGGTCCCAGGCCTTGCTGGGCCAGACTGATGG 1993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCAGGAATTCGCCCGGGTTACTTCCGGAGGTCCCAGGCCTTGCTGGGCCAGACTGATGC 120

Qy   1994 AGGTGCCCATAAAAGACACCGACAGCTGCTGCAGTACACCAAGGCTCCCATCCAGGA 2053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGTGCCCAGGAAAGCACCGACAGCTGCTGCAGTACACCAAGGCTCCCATCCAGGA 180

Qy   2054 GTCAATTTTCAGCTCAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGCTTCCTGGCCCT 2113
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTCGCTCTTCAGCTCAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGCTTCCTGGCCCT 240

Qy   2114 GATGCGGTTTATGGGTGATTATGTCAAGCCCCCGGGCAAGGATGAGATGGATCTGCTCTA 2173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GATGCGGTTTATGGGTGATTATGTCAAGCCCCCGGGCAAGGATGAGATGGATCTGCTCTA 300

Qy   2174 TGAACTGCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTAT 2233
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGAACTGCTGAAGCTGTGCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTAT 360

Qy   2234 CAAGCAGGTCACGGGACAGCCCCCGGCT ---------- ------- ------- --- 2293
           |||||||||||| ||||||||||||||
Db    361 CAAGCAGGTCACGGGACAGCCCCCGGCT ---------- ------- ------- --- 387

Qy   2294 GGTTGTCACAGATTTCTT [illegible] TGATGCCCTACCTGACCAAGTTTCT 2353
                                ||||||||||||||||||||||||||||||
Db    388 ----------------------------------GCTGATGCCCTACCTGACCAAGTTTCT 414

Qy   2354 GCAAACTTCAGCCCCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCAGCTCCAGCGCAC 2413
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    415 GCAGGATTCAGCCCCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCAGCTCCAGCGCAC 474

Qy   2414 AGTCAAATATGGGGGGCGCCGGCGGATGCCCCCACCGGGTGAAATGAAGGCTTTCCTGAA 2473
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    475 AGTCAAATATGGGGGGCGCCGGCGGATGCCCCCACCGGGTGAAATGAAGGCTTTCCTGAA 534

Qy   2474 AGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCCGGGCGTGTGGATTATAGGACGAA 2533
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    535 AGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCCGGGCGTGTGGATTATAGGACGAA 594

Qy   2534 TATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCCCCAAATGGGTAT 2593
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    595 TATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCCCCAAATGGGTAT 654

Qy   2594 CAGGGAGGCTCAGGAAGTGCAGGAATTCGCCCTGTTCTTCATCAAAGACAAGCAGCCAGCT 2653
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    655 CAGGGAGGCTCAGGAAGTGCAGGAATTCGCCCTGTTCTTCATCAAAGACAAGCAGCCAGCT 714

Qy   2654 GGTGCGGCCCCTGCAGCCCGCCGAATACCTCAACAGCGTGTACTGTGACCAGGACCTGAG 2713
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    715 GGTGCGG-CCCTGCAGCCCGCCGAATACCTCAACAGCGTGTACTGTGACCAGGACCTGAG 773

Qy   2714 CCTGCACAGGGGCGGCTCCACTGGGAGA-CCCCACTGCACTTCGA-TAACTCCACCTACA 2771
           |||||||||| | || |||||||||||| |||||||||||||||| ||||||||||||
Db    774 CCTGCACAGCCGGGGGTCCACTGGGAGACCCCCACTGCACTTTGATTAACTCCACCTACA 833

Qy   2772 TCA 2774
           |||
Db    834 TTA 836


RESULT 6
BM923763
LOCUS       BM923763                1015 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6709671 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5759880
            5', mRNA sequence.
ACCESSION   BM923763
VERSION     BM923763.1  GI:19774142
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1015)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12806  row: a  column: 01
            High quality sequence start: 222
            High quality sequence stop: 720.
FEATURES             Location/Qualifiers
     source          1. .1015
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5759880"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI, Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 4 colons, age 56 y, male; 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 Kb,
                     insert size range 1-3 kb.  Library is normalized and
```

```
                    enriched for full length clones and was constructed by C
                    Gruber  (Invitrogen)  Research Genetics tracking code
                    023. Note: this is a NIH_MGC Library."
BASE COUNT      218 a    317 c    294 g    173 t      3 others
ORIGIN

  Query Match             18.6%;  Score 699.2;  DB 14;  Length 1015;
  Best Local Similarity   92.5%;  Pred. No. 1e-95;
  Matches  760;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    3;

Qy    2192 GCTTCGGGGCCGAACAGTGGAGTCGAGGGTGGAGGTTCCTCAGGCTTCTCACAGGGCTTCTT 2311
           |||| |||| | || |||||| ||     |||||||  |||||||||  |||||
Db      49 GCCTAGGAGGAAGAACTACACTCGACGCTCTGCCTTCCTCACGCTTCTCACAGGGTTCTT 168

Qy    2312 CCCTCTCCTCAGCCAGGTCATCGTCTACCTGACCAAAGTTCTCTCAGATTCA 2465
           |||| || | |||| |||||| |  |||||||| |||||||||  |||| |||||||
Db     169 CCTTCCAACCACCACTGTGATGCTGTTGCTGACCAACTTTCTCTGCAGCCTTCACGGCCCCA 228

Qy    2466 GCTAAGAACTGGCCTGGAGTAGCCAGGAGTACTCTCAGCACACAGTCAAATATGTGGGGGC 2625
           || ||||   ||||| ||||||||||||||||||| ||||||||||||||||||||
Db     229 TCTGACAGCTGTGCCGAATACCCAGGAGCACTTCTCAGCACACAGTCAAATATGTGGGGGC 288

Qy    2626 TCTGGCGATGCCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGATTCGCC 2785
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     289 TCTGGCGATGCCCCCACCGGCTGAAATGAAGGCTTTCCTGTAAGGACAAGCGATTCGCC 348

Qy    2786 TCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAAATATCCAGACTTTCACAG 2945
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     349 TCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAAATATCCAGACTTTCACAG 408

Qy    2946 TATAAGTAAAATTTACTATTTTTCCCCCAAATGGTTATTGCAGGAGCTTCAGGAAG 3105
           || ||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     409 TACAGCAGAAGTGTAGGAGGAGCTGTGCCGGCAAATGGGTATCAGGAGCCTCAGGAAG 468

Qy    3106 TCTAACAATTCCGCCCTCTTCCTCATAAAAGAAGAGCCAGCTGGTGGGGCCCTGCAGC 3265
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     469 TCTAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGGGGCCCCTGCAGC 528

Qy    3266 TCTAATGGAATAAGTCAACAGCCTGCTATGGACAAGGAGCTGAGCCTGCACAG-GGGTGGC 3424
           ||| |||||| |||||||||||||||||||||||||||||||||||||| ||||||||
Db     529 TCCACTGGAATACCTCAACAGCCTGCTATGCACAAGGAGCTGAGCCTGCACAGCGGGTGGC 588

Qy    3425 TCTATTGGGAGACCCCACTGCACTTCGATAACTTCAACTACATCAGCAACCCACTACAGCC 3584
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     589 TCTACTGGGAGACCCCACTGCACTTCGATAACTTCAACTACATCAGCAACCCACTACAGCC 648

Qy    3585 TCTCTTCTGGGACTACCTTCAGGGGAAAGTCGCCAGTCAGGGCAAGGCAGACGTGCAGC 3744
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     649 ACTCTTCTGGGACTACCTTCAGGGGAAAGTCGCCAGTCAGGGCAAGGCAGACGTGCAGC 708

Qy    3745 TCTCTAGCTGGGCCGCCTGCAGCACCTCAGCAAGGGCCAACAGGAATACCCCCTCAGGGC 3904
           ||  ||| ||||||||||||||||||||||||||||||| | ||||||||||||||||
Db     709 TCTCTAGCTGGGCCGCCTGCAGCACCTCAGCAAGGGCCACAGGAATACCCCCTCAGGGC 768

Qy    3905 ACCACTCTGCTAGCTTAGCTGCCAAAGCAGTTGCAAGGGCAGGTGAACACGGGCCTCCATCA 4064
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     769 ACCACTCTGCTAGCTTAGCTGCCAAAGCAGTTGCAAGGGCAGGTGAACACGGGCCTCCATCA 828

Qy    4065 GTAAACTTCATGGGTCAGGAGCTGAAGACGGCT-GGAAGGACACAGCCCCCAGGAAGCACAG 4223
           ||| |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     829 GTAAACTTCATGGGTCAGGAGCTGAAGACGGCTTGGAAGGACACAGCCCCCAGGAAGCACAG 888

Qy    4224 ATCAGCTTCATTGAAGGCCATGAGCCAGCTGGCCCTCTTCGGCTAACACCG 4373
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     889 ATCAGCTTCATTGAAGGCCATGAGCCAGCTGGCCCTCTTCGGCTAACACCG 938


RESULT 2
BI766411
LOCUS       BI766411                 908 bp    mRNA    linear   EST 08-DEC-01
DEFINITION  yo36b06.s1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201966 5',
            mRNA sequence.
ACCESSION   BI766411
VERSION     BI766411.1  GI:15778989
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11506  row: b  column: 15
            High quality sequence stop: 852.
FEATURES             Location/Qualifiers
     source          1..908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201966"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source:
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20 22 week male spleens.  Library is oligo dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning)  Average insert size 1.4 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen)  Research Genetics tracking  code 023. Note:
                     this is a NIH_MGC Library."
BASE COUNT      209 a    287 c    272 g    140 t
ORIGIN

  Query Match             18.2%;  Score 687.4;  DB 14;  Length 908;
  Best Local Similarity   91.2%;  Pred. No. 4.1e-84;
  Matches  828;  Conservative    0;  Mismatches   84;  Indels    5;  Gaps    5;

Qy    2234 CAAGCAGGTCACAGGACACCCCGGCCCGAAATACTGCACTCCTCTCATC 2383
           ||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1 CAAGCAGGTCACGGGACACCCCGGCCCGAAATACTGCACTCCTCTCATC 50

Qy    2294 CCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGATCCCCAACTGTTCT 2443
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCTTCTCACAGGCTTCTTCCCCCCGTCGACCAGGCTGATCCCCAACTGTTCT 110

Qy    2354 GCAGGATTCAGGCCCCAGCCAAGAGCTGGCCCGGAGCAGCAAGACCTCCAGGCAC 2413
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCAGGATTCAGGCCCCAGCCAAGAGCTGGCCCGGAGCAGCAAGACCTCCAGGCAC 180

Qy    2414 AGTCAAATATGGGGGGCGCCGGCGGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAA 2473
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGTCAAATATGGGGGGCGCCGGCGGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAA 240

Qy    2474 AGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGAA 2633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 AGGACAAGCGATTCGCCTGCTTCTTATTCACCTGCCGGGGGGGTGTGGATTATAGGACGAA 299

Qy    2634 TATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTAT 2793
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 TATCCAGACTTTCACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCCGGCAAATGGGTAT 359

Qy    2794 ATCAGGAGCCTCAGGAAGTGCAGGAATTCCCCCTCTTCCTCATCAAAGAGAAGAGCCA 2953
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 CACGGAGCCTCAGGAAGTGCAGGAATTCCCCCTCTTCCTCATCAAAGAGAAGAGCCA 419
```

```
Qy   2647 ----------------------------  ------  ----GCCAGCTGGTGCGGCCCC 2664
                                                  ||||||||||||||||||||
Db    429 TGGGGACCGAGCCTTACTGGTGGGGTCACTCCCTGCCCAGGGGCCAGCTGGTGCGGCCCC 479

Qy   2665 TGCAGCCCGGCGAATACTTAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAG C 2723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    480 TGCAGCCCGGCGAATACTTAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCC 539

Qy   2724 GGGGGTGCACTGAGAGACCCACTGCACTTCGATAACTCCACCTACATCAGCACCCACT 2783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GGGGGTGCACTGAGAGACCCACTGCACTTCGATAACTCCACCTACATCAGCACCCACT 599

Qy   2784 ACAGCCAGGTGCTGTGGAACTACTTCAGGGGAAGCTGCCAGTCAGCGCCAAGGCAGACG 2843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 ACAGCCAGGTGCTGTGGAACTACTTCAGGGGAAGCTGCCAGTCAGCGCCAAGGCAGACG 659

Qy   2844 CGCAGTTGGCAGGCTGGCGGGCCTGCAGCACCTCAGCAAGGCCAACAGGAATACCCCCT 2903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 CGCAGTTGGCAGGCTGGCGGGCCTGCAGCACCTCAGCAAGGCCAACAGGAATACCCCCT 719

Qy   2904 GAGGGAGGACTTTCTAGTTAAGTGGCCAAAGCAGCT-GCAAGGGCAGGTGAACAGGGC 2962
          |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||| |
Db    720 CAGGGCAGCAGCTGCTAGTTAACGTGCCAAAGCAGCTGGAAACGGCAGGTGAACACGGGC 779

Qy   2963 TCCATCAAGAAGCTGATGGG--TCAGGAGCTGAGAGGCTGGAA-GGACACAGCCCCCAG 3019
          |||||||||||||||||||  ||||||||| |||||||| ||||||||||||||||||
Db    780 TCCATCAAGAAGCTGATGGGGTCAGGAGCTGAGAAAGCTGGAAGGGACACAGCCCCCAG 839

Qy   3020 GAAGCACAGAT-CAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGT 3078
          |||||||||||| |||||||||||||||||||||||||||||||||||| | | | ||
Db    840 GAAGCACAGATCCAGCTTCATTGAGGCCATGAGCCAGCTGCCCATCTACGGGTAAACGG 899

Qy   3079 CTATGGGG 3086
          ||||||||
Db    900 CTATGGGG 907


RESULT 8
BM977966/c
LOCUS       BM977966                689 bp    mRNA    linear   EST 21-MAR-2002
DEFINITION  UI-CF-EC1-aec-d-23-0-UI s1 UI-CF-EC1 Homo sapiens cDNA clone
            UI-CF-EC1-aec-d-23-0-UI 3', mRNA sequence.
ACCESSION   BM977966
VERSION     BM977966.1  GI:19596016
KEYWORDS    EST
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4966
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..689
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EC1-aec-d-23-0-UI"
                     /clone_lib="UI-CF-EC1"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: Lung; Vector: pT7T3 Pac (Pharmacia) with a
                     modified polylinker, Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EC1 is a normalized cDNA library containing the
                     following tissue(s): Normal lung from adult and from fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC
                     TAG_LIB=UI-CF-EC1
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_SEQ=AAGTGCTTAC"
BASE COUNT      105 a    192 c    225 g    166 t      1 others
ORIGIN

  Query Match          18.1%;  Score 683.2;  DB 14;  Length 689;
  Best Local Similarity   99.4%;  Pred. No. 1.8e-83;
  Matches  685;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   3083 GGGGTGCTGCGAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGC 3142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    689 GGGGTGCTGCGAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGC 630

Qy   3143 CAGCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAG 3202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 CAGCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAG 570

Qy   3203 AGCCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCTGGCCTGGAA 3262
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    569 AGCCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCTGGCCTGGAA 510

Qy   3263 GTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGCTGCCACAGGCCCAG 3322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509 GTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGCTGCCACAGGCCCAG 450

Qy   3323 GAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCCTCTTGCACTGAGTGGCCC 3382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 GAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCCTCTTGCACTGAGTGGCCC 390

Qy   3383 AGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCCTCCCCCGCCCCA 3442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    389 AGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCCTCCCCCGCCCCA 330

Qy   3443 AGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCA 3502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 AGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCA 270

Qy   3503 GCACAGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGACCTGCTATCAGGTCACCAGC 3562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 GCACAGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGACCTGCTATCAGGTCACCAGC 210

Qy   3563 AGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGA 3622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 AGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGA 150

Qy   3623 ACTCGGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGACCCCAGGCT 3682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 ACTCGGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGACCCCAGGCT 90

Qy   3683 CCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGGTGCTGAGGAAATA 3742
```

```
Db      89 AAGCCAGGGCCCACATTAGCAAGCCCAGGCATGGGAGAAACAGCTGCTGAGAAATA 30

Qy    3743 AAACTGCCTAAAAAAAAAAAAAAAAAAA 3771
           ||| |||||  | ||||||||||||||||
Db      29 AAACTGCCTGGAGAAAAAAAAAAAAAAAA 1


RESULT 9
BM720779
LOCUS       BM720779                 728 bp    mRNA    linear   EST 01-MAR-2002
DEFINITION  UI-E-EO1-aib-g-07-0-UI.s1 UI-E-EO1 Homo sapiens cDNA clone
            UI-E-EO1-aib-g-07-0-UI 5', mRNA sequence.
ACCESSION   BM720779
VERSION     BM720779.1  GI:19040260
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Gene Discovery, and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares-blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..728
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EO1-aib-g-07-0-UI"
                     /clone_lib="UI-E-EO1"
                     /tissue_type="fetal eye"
                     /dev_stage="fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker, Site_1: EcoR I, Site_2: Not I,
                     UI-E-EO1 is a normalized cDNA library containing the
                     following tissue(s): fetal eye. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     CGCGTATACC. This library was created for the program, Gene
                     Discovery in the Visual System, supported by National Eye
                     Institute (NEI)."
BASE COUNT      142 a    210 c    222 g    132 t      2 others
ORIGIN

  Query Match          18.2%;  Score 681.2;  DB 14;  Length 728;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-83;
  Matches  716;  Conservative    0;  Mismatches    4;  Indels    3;  Gaps    3;

Qy    2018 AAAATAAACAGCCT CGTGCACTATACTAAAGTTTTCATTCAGGAGTCGTTTTTATGTT 2069

Db       1 GAACAGGAGAAGTTGGGGGGAGAGACAAAAAGGGAGTCAAAGAGGGTCTTAGTTT 56

Qy    2069 AAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGTTGCTaaaCCGAGTCGTTTTATGGG 2128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57 CAGTGATGATGTGAGCAAGCTGGCTGTAGCCAGTTGCTaaaCCGAGTCGTTTTATGGG 116

Qy    2129 TGACCAGTCCAAGCCCCGGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAAGCT 2188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117 TGACCAGTCCAAGCCCCGGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAAGCT 176

Qy    2189 GTGCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGCTCAGCAG 2248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 GTGCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGCTCAGCAG 236

Qy    2249 ACACCCCTGGGCCGGAACACTGCACTCCAAGCTCCAAGCTCCCCCCCCCCCCCCCCCCCC 2308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237 ACACCCCTGGGCCGGAACACTGCACTCCAAGCTGGAGCTCCTCCAGCAGCTTCTCAGGTT 296

Qy    2309 CTCCCCCCCCCTCAGCAGCTGATGCCCTACCTGACCAAGCTCTTCTTAGAATTCAGCCCC 2368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 CTCCCCCCCCTCAGCAGCTGATGCCCTACCTGACCAAGCTTTCTTCAGGATTCAGGCCC 356

Qy    2369 CAGCCAAGAGCTGGCCCTGGAGCAGCCAGGAGCACCCCAGCCCCACAGTCAAATATCCCCC 2428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     357 CAGCCAAGAGCTGGCCCTGGAGCAGCCAGGAGCACCCCAGCCCCACAGTCAAATATCCCCC 416

Qy    2429 GCGCCGGCGGATGCCCCCACCGGCTGAAATGAAGGCTTTCCTGAAAGCGCAAGCGATTCC 2488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     417 GCGCCGGCGGATGCCCCCACCGGCTGAAATGAAGGCTTTCCTGAAAGCGCAAGCGATTCC 476

Qy    2489 CCTGCTTCTTATTCACCTGCCGGGGGGGTTTGCATTATAGGAACCAAATATCCAGACTTTCAC 2548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     477 CCTGCTTCTTATTCACCTGCCGGGGGGGTTTGCATTATAGGAACCAAATATCCAGACTTTCAC 536

Qy    2549 AGTAGCAGCAGAAGTTTACGAGTCTTTTTTCTCAAATTTTGATTACTGCAGCAGCTCCAGA 2608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     537 AGTAGCAGCAGAAGTTTACGAGTCTTTTTTCTCAAATTTTGATTACTGCAGCAGCTCCAGA 596

Qy    2609 AGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCTACTCCTGGTTCTCCTCCA 2668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     597 AGTGCAGGAATTCGCCCTCTTCCTCATCAAAGAGAAGAGCTACTCCTGGTTCTCCTCCA 656

Qy    2669 GCCCGCCGAATACTTCAACAGCGTGGTACTGGA CAGGACGTGCAG CTGCACAGCGGCGG 2728
           ||||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||
Db     657 GCCCGCCGAATACTTCAACAGCGTGGTACTGGA CAGGACGTGCAG CTGCACAGCGGCGG 716

Qy    2729 CTC 2731
           |||
Db     717 CTC 719


RESULT 10
BI769659
LOCUS       BI769659                 661 bp    mRNA    linear   EST 15-OCT-2001
DEFINITION  602855559F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5,00 ,
            mRNA sequence.
ACCESSION   BI769659
VERSION     BI769659.1  GI:15761237
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-email@nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I M A G E  Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM11514  row: m  column: 13
          High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1   661
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5205300"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      132 a    239 c    176 g    114 t
ORIGIN

  Query Match           16.0%;  Score 606.4;  DB 13;  Length 661;
  Best Local Similarity   97.2%;  Pred. No. 4.6e-73;
  Matches  653;  Conservative    0,  Mismatches    6;  Indels   13;  Gaps    3;

Qy     517 CCAAGCCGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGCATGTACCAGAGCCGCC 576
           || |   | ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       2 CCCAAGCCCCACAACCCAGCAGGAAATTGGCAACATCATCCGCATGTACCAGAGCCGCC 60

Qy     577 CGGGCCCCGTGCCTGTGCCGTGCAGCCATCC     ...     AGCTTTCCTGAGGAAAA 626
           ||||||||||||||||||||||||||||||||                ||||||||||||||||
Db      61 CGGGCCCCGTGCCTGTGCCGTGCAGCCATCC                 AGCTTTCCTGAGGAAAA 109

Qy     627 TCGACCCCAAGGAAGAGGTCTTGGCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGC 686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     110 TCGACCCCAAGGAAGAGGTCTTGGCCAAGCTGGGTATCAACGGTGCCCACTCGTCCCCGC 169

Qy     687 CGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCCGCCAGCTGTGGCTCCTCGACCCAAGG 746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     170 CGATGCTGTCCCCCAGCCCAGGAAAGGGCCCCCGCCAGCTGTGGCTCCTCGACCCAAGG 229

Qy     747 CCCCGTACAGCTTGGGCCCTCTAGCTGCATCAAGGAAAAGCAGGGGCCCCTTCTGGACC 806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     230 CCCCGTACAGCTTGGGCCCTCTAGCTGCATCAAGGAAAAGCAGGGGCCCCTTCTGGACC 289

Qy     807 TGTTTGTCCAGAAGCTGCCTATTGCCACACACCCCACCTCCACCAGCGCCACCACTGC 866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     290 TGTTTGTCCAGAAGCTGCCTATTGCCACACACCCCACCTCCACCAGCGCCACCACTGC 349

Qy     867 TCTCCCCAAGAAGCCCAGGGACCGTTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGG 926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     350 CTCTGCCCAGGACCCAGGGACCGTTTCAGCAGAGCGTCGTTGCTTGACACAGCCCGTGG 409

Qy     927 AGGACCAGGGGGTCTCCACCAGTACTACTGCGCCCTCTGGCAGCGTGTGC-TTCTCCTAC 985
           ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     410 AGGACCAGGGGGTCTCCACCAGTACTACTGCGCCCTCTGGCAGCGTGTGCTTTCTCCTAC 469

Qy     986 AGCGGCACCGCCTTGGAAGTTGTTCCTACCCAAGGAGGTGTTCTACCCACGGGAGAACTTC 1045
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     470 ACCGGCACCGCCTTGGAAGTTGTTCCTACCCAAGGAGGTGTACTACCCACGGGAGAACTTC 529

Qy    1046 AGCCATCCCTACTACCTGAAGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAG 1105
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     530 AGCCATCCCTACTACCTGAAGCTCCTCTGTGAGCAGATCCTACGGGACACCTTCTCCGAG 589

Qy    1106 TCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGGCTTG 1165
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     590 TCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGGCTTG 649

Qy    1166 GAGGTGGACCTG 1187
           ||||||||||||
Db     650 GAGGTGGACCTG 661


RESULT 11
BM791149
LOCUS       BM791149                 603 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  K-EST0071111 S21SNU520 Homo sapiens cDNA clone S21SNU520-19-F03 5',
            mRNA sequence.
ACCESSION   BM791149
VERSION     BM791149.1  GI:19139381
KEYWORDS    EST
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 19  row: F  column: 03
            High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1..603
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="S21SNU520-19-F03"
                     /clone_lib="S21SNU520"
                     /sex="F"
                     /tissue_type="Stomach"
                     /cell_type="Floating aggregates"
                     /cell_line="SNU-520"
                     /lab_host="Top10F'"
                     /note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
                     Site_2: NotI, The poly (A)+ RNA was bpl phosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including EcoR
                     I site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized from oligo dT-selected mRNA by
                     priming with dT-tailed vector. The dT-tailed vector was
                     adjusted to have about 60nt. The cDNA vector was
                     circularized with E. coli DNA ligase after digestion of
                     EcoRI which site is also included in vector. An RNA strand
                     converted to a DNA strand by Okayama Berg method. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
BASE COUNT      109 a    186 c    187 g    121 t
ORIGIN

  Query Match           16.0%;  Score 603;  DB 14;  Length 603;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-72;
  Matches  603;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1114 AGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGGCT 1173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAAGACCTGCTGGGAGGCT 60

Qy    1174 TGGAGGTGGACCTGGATTCTTCAACACCAGGAAGACAGGTCAAGAAGCGCATTTCT 1230
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     61 [illegible] 120

Qy   1234 TGGCCGCTCGGGACAACTGGGCCAATTACTTCTCCGG[illegible]GAGA 1293
          [illegible]
Db    121 TGGCCGCTCGGGACAACTGGGCCAATTACTTCTCCGGCTTCTTCTTGTCTCGGGCGAGA 180

Qy   1294 GTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCCACCGTGGGCTGGACTGCTAAGGTGA 1353
          [illegible]
Db    181 [illegible] 240

Qy   1354 [illegible] 1413
          [illegible]
Db    241 [illegible] 300

Qy   1414 AGG[illegible]CCTGGAGCTGTCACTGAAGAGCGAGC 1473
          [illegible]
Db    301 [illegible] 360

Qy   1474 [illegible] 1533
          [illegible]
Db    361 [illegible] 420

Qy   1534 [illegible] 1593
          [illegible]
Db    421 [illegible] 480

Qy   1594 [illegible] 1653
          [illegible]
Db    481 [illegible] 540

Qy   1654 [illegible]GACATAG 1713
          [illegible]
Db    541 AGCCAGGCTGGCAGTTTGGCTCT[illegible]GACATAG 600

Qy   1714 TG[illegible] 1716
          ||
Db    601 TG[illegible] 603
```

```
RESULT 12
AW753302
LOCUS       AW753302                626 bp    mRNA    linear   EST 28-APR-2000
DEFINITION  [illegible] Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW753302
VERSION     AW753302.1  GI:7668234
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   [illegible] Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, [illegible], Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=[illegible]
            [illegible] 011 b05&t3=1999 [illegible]
            Seq primer: puc 18 forward
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1..626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0254"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. [illegible]
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     [illegible] the puc 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      149 a    195 c    184 g     97 t      1 others
ORIGIN

  Query Match             15.8%;  Score 595.8;  DB 1[illegible];  Length 626;
  Best Local Similarity   99.2%;  Pred. No. 1.3e-71;
  Matches  [illegible];  Conservative    0;  Mismatches    [illegible];  Indels    [illegible];  Gaps    [illegible];
```

```
Qy   2340 [illegible] 2399
          [illegible]
Db      4 CTGACCAAGTTTCTGCCAGGATTCAGGCCCCAGCCAAGAGCTGG[illegible]GAGCAGGAG 63

Qy   2400 [illegible]GAAAT 2459
          [illegible]
Db     63 [illegible]GAAAT 122

Qy   2460 [illegible] 2519
          [illegible]
Db    123 [illegible] 182

Qy   2520 GATTATAGGACGAATATCCAGACTTTCACAGTAGCAGCAGAAG[illegible] 2579
          [illegible]
Db    183 GATTATAGGACGAATATCCAGACTTT[illegible] 242

Qy   2580 CGGCAAATGGGTATCACGGAGCCTCAGGAAGTGCAGGAATT[illegible]TCATCAAA 2639
          [illegible]
Db    243 CGGCAAATGGGTATCACGGAGCCTCAGGAAGTGCAGGAATT[illegible]TCATCAAA 302

Qy   2640 GAGAA[illegible] 2699
          [illegible]
Db    303 GAGAAGAG[illegible] 362

Qy   2700 GACCAGGACGTGAGCCTGCACAG[illegible]GAT 2758
          [illegible]
Db    363 GACCAGGACGTGAGCCTGCACAG[illegible]GAT 422

Qy   2759 AACTCCACCTACATCAGCACCCACTACAGCCAGGTGCTGTGGGACTACCTTCAGGGGAAG 2818
          [illegible]
Db    423 AACT[illegible]TTCAGGGGAAG 482

Qy   2819 [illegible]AGCTCGCCAGGCTG[illegible] 2878
          [illegible]
Db    483 CTGCCAGTCAGGGCAAGGCAGACGGCAGCTCGCCAGGCTGG[illegible] 542

Qy   2879 AGCAAGGCCAACAGGAATACCCCCTCAGGGCAGGACCTGCTAGCTTACGTGCCAAAGCAG 2938
          [illegible]
Db    543 AGCAAGGCCAACAGGAATACCCCCTCAGGACAGGACCTGCTAGCTTACGTGCCANAGCAG 602

Qy   2939 CTGCAACGGCAGGTGAACACGGCC 2962
          [illegible]
Db    603 CTGCAACGGCAGGTGAACACGGGC 626
```

```
RESULT 13
BG949021/c
LOCUS       BG949021                623 bp    mRNA    linear   EST 12-JUN-[illegible]
DEFINITION  [illegible] HP0027 [illegible] Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG949021
VERSION     BG949021.1  GI:[illegible]
```

```
KEYWORDS    EST.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Dias Neto,E., Garcia Correa,P., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getHuml2.pl?l1 QV1&l2 QV1-HB0037-
            020201.599.e08&t2 2001-02-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 582
FEATURES             Location/Qualifiers
     source          1..623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HB0037"
                     /dev_stage="Adult"
                     /note="Organ: brain_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       93 a    171 c    218 g    141 t
ORIGIN

  Query Match           15.3%;  Score 580;  DB 13;  Length 623;
  Best Local Similarity   99.2%;  Pred. No. 1.9e-69;
  Matches  605;  Conservative    0;  Mismatches    0;  Indels    5;  Gaps    2;

Qy    2623 CCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGCGGCCCCTGCAGCCCGCCGAATACC 2682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     623 CCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGCGGCCCCTGCAGCCCGCCGAATACC 564

Qy    2683 TCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAG-CGGCGGCTCCACTGGGAGA 2741
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     563 TCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCCGGCGGCTCCACTGGGAGA 504

Qy    2742 CCCCAATTGCACTTCGATAACTCCACCTACATCAGCACCCACTACAGCCAGGTGCTGTGGG 2801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     503 CCCCAATTGCACTTCGATAACTCCACCTACATCAGCACCCACTACAGCCAGGTGCTGTGGG 444

Qy    2802 ACTACTTTCAGGGAAGCTGCCAGTCAGCGCCAAGGCAGACGCGCAGCTCGCCAGGCTGG 2861
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     443 ACTACTTTCAGGGAAGCTGCCAGTCAGCGCCAAGGCAGACGCGCAGCTCGCCAGGCTGG 384

Qy    2862 CCGCCCTGCAGCACCTCAGCAAGGCCAACAGGAATACCCCCTCAGGGCAGGACCTGCTAG 2921
           ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db     383 CCGCCCTGCAGCACCTCAGCAAGGCCAACAGGAATACCCCCTCAG----GGACCTGCTAG 328

Qy    2922 CTTACGTGCCAAAGCAGCTGCAACGGCAGGTGAACACGGCCTCCATCAAGAACCTGATGG 2981
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 CTTACGTGCCAAAGCAGCTGCAACGGCAGGTGAACACGGCCTCCATCAAGAACCTGATGG 268

Qy    2982 GTCAGGAGCTGAGACGGCTGGAAGGACACAGCCCCAGGAAGCACAGATCAGTTCATTG 3041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 GTCAGGAGCTGAGACGGCTGGAAGGACACAGCCCCAGGAAGCACAGATCAGTTCATTG 208

Qy    3042 AGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGTCTATGGGGTGCTGCGAGTGAGCA 3101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 AGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGTCTACGGGGTGCTGCGAGTGAGCA 148

Qy    3102 TGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACGGCCAGCATCTCATCCTCATGG 3161
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147 TGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACGGCCAGCATCTCATCCTCATGG 88

Qy    3162 ACCCCAGCTCCCAGAGCCTGTACTGCGCATTGCCCTGAAGAGCCTGCAGGGGCTCCACC 3221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87 ACCCCAGCTCCCAGAGCCTGTACTGCGCATTGCCCTGAAGAGCCTGCAGGGGCTCCACC 28

Qy    3222 TGCTAAGCCC 3231
           ||||||||||
Db      27 TGCTAAGCCC 18


RESULT 14
AV683102
LOCUS       AV683102                676 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  AV683102 GKC Homo sapiens cDNA clone GKCEBF11 5', mRNA sequence.
ACCESSION   AV683102
VERSION     AV683102.1  GI:10284965
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,X., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex 45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GKCEBF11"
                     /clone_lib="GKC"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      159 a    227 c    183 g    115 t      1 others
ORIGIN

  Query Match           15.2%;  Score 573.4;  DB 10;  Length 676;
  Best Local Similarity   94.7%;  Pred. No. 1.3e-68;
  Matches  638;  Conservative    0;  Mismatches   27;  Indels    9;  Gaps    4;

Qy    2664 CTGCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAG 2723
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAG 60
```

```
Qy    2724 CGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCAGCACCCAC 2783
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCAGCACCCAC 120

Qy    2784 TACAGCCAGGTGCTGTGGGACTACCTTCAGGGAAGCTGCCAGTCAGCGCCAAGGCAGAC 2843
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TACAGCCAGGTGCTGTGGGACTACCTTCAGGGAAGCTGCCAGTCAGCGCCAAGGCAGAC 180

Qy    2844 CCCCAGCTCCCCAGGCTGGCCCCCTGCAGCACCTCAGCAAGCCAACAGGAATACCCCC 2903
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCCCAGCTCCCCAGGCTGGCCCCCTGCAGCACCTCAGCAAGCCAACAGGAATACCCCC 240

Qy    2904 TCACCCCAGGACCTCCTAAGTTACGTGCCAAAGCAGCTGCAACGGCAGGTGAACACGGCC 2963
           ||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCAG    GGACCTCCTAGCTTACGTGCCAAAGCAGCTGCAACGGCAGGTGAACACGGCC 296

Qy    2964 TCCATCAAGAACCTCATCGGGTCAGGAGCTGAGACGGCTGGAAGGACACAGCCCCCAGGAA 3023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 TCCATCAAGAACCTCATCGGGTCAGGAGCTGAGACGGCTGGAAGGACACAGCCCCCAGGAA 356

Qy    3024 GCACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGTCTAT 3083
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     357 GCACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGTCTAT 416

Qy    3084 GGGGTGCTGCGAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGC 3143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     417 GGGGTGCTGCGAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGC 476

Qy    3144 CAGTATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAG 3203
           ||||||||||  ||||||||||       || ||||||||||  |||| ||||||||||
Db     477 CAGCATCTCATACTCATGGACCCCTAGCTTCAAAGCCTGTACTGTCGCAATGCCCTGAAG 536

Qy    3204 AGCCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGGCCTGGAA 3263
           |||||||| |||||  ||||||||||||    |  |   |||||  ||||||||||||  ||
Db     537 AGCCTGCAACGGCT-CACCTGCTAAGCCTCTGGAGGAGAAAGGGNCCCCCTGGCCTGTAA 595

Qy    3264 CTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGCTGCCACAGGCCCAG 3323
           ||   ||||||||||||||||||||  ||||||||||||||||||  ||||||||||||
Db     596 GT-  CTATGGCTCAGCTGACAACCCTCAGACCATCTGGTTTGAGCTTGCACAGGCCCAG 652

Qy    3324 GAGCTGCTATACAC 3336
           |||||| ||||| |
Db     653 GAGCTGGTATACCC 666


RESULT 15
AI825877/c
LOCUS       AI825877                 567 bp    mRNA    linear   EST 24-AUG-1999
DEFINITION  td19g01.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2076144 3',
            mRNA sequence.
ACCESSION   AI825877
VERSION     AI825877.1  GI:5446548
KEYWORDS    EST.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaphs-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael P. Emmert-Buck, M.D.
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1..567
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2076144"
                     /clone_lib "NCI_CGAP_Co16"
                     /tissue_type="colon tumor, RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pCMT3D Pac (Pharmacia) with a
                     modified polylinker; Site 1: Not I; Site 2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Co16 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1057416-1061255, and 1144594-1149251).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       96 a    152 c    186 g    133 t
ORIGIN

  Query Match             14.9%;  Score 564;  DB 9;  Length 567;
  Best Local Similarity   100.0%;  Prod. No. 2.7e-67;
  Matches  564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3188 CGCATTGCCCTGAAGAGCCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGG 3247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     567 CGCATTGCCCTGAAGAGCCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGG 508

Qy    3248 CCCCCTGGCCTGGAAGTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAG 3307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 CCCCCTGGCCTGGAAGTCAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAG 448

Qy    3308 CTGCCACAGGCCCAGGAGCTGCTATACACCACTGTCTTCCTGGATAGACAGCAGTGGCCTCT 3367
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 CTGCCACAGGCCCAGGAGCTGCTATACACCACTGTCTTCCTGGATAGACAGCAGTGGCCTCT 388

Qy    3368 TGCACTGAGTGGCCCAGCATCAACTGACAGGAGTGCAGGCCCGGGGAGAGAAGAGGATGAG 3427
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 TGCACTGAGTGGCCCAGCATCAACTGACAGGAGTGCAGGCCCGGGGAGAGAAGAGGATGAG 328

Qy    3428 GCCTCCCCCGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTT 3487
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 GCCTCCCCCGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTT 268

Qy    3488 CTAGAACCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCCATGAGGAGGCTGC 3547
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 CTAGAACCTGCCTCAGCACAGCCCAGCCGGCCCACATGCAGGCCATGAGGAGGCTGC 208

Qy    3548 TATCACGTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCTGGGCCA 3607
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 TATCACGTCACCAGCAGGCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCTGGGCCA 148

Qy    3608 AAGGGGCTGCAGGAACTGGCTGGGGCACCTGAGGTTCCCAGTCTGAGGGACATGCCC 3667
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147 AAGGGGCTGCAGGAACTGGCTGGGGCACCTGAGGTTCCCAGTCTGAGGGACATGCCC 88

Qy    3668 ACCCGACCCCAGGCTCCGGCCAGGCCCACATTAGCACAAGCCCAGCATGGACAAACA 3727
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87 ACCCGACCCCAGGCTCCGGCCAGGCCCACATTAGCACAAGCCCAGCATGGACAAACA 28

Qy    3728 GCTGCTGAGGAAATAAAACTCCCT 3751
           ||||||||||||||||||||||||
Db      27 GCTGCTGAGGAAATAAAACTCCCT 4


Search completed: January 10, 2003, 01:55:58
Job time : 4540.17 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:10:54 ; Search time 107.379 Seconds
(without alignments)
10795.718 Million cell updates/sec

Title: US-09-803-126-7
Perfect score: 3780
Sequence: 1 cggcagjagcaggattgggc ... aaaaaaaaaaaaaaaaaaa 3780

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
C 1	73.4	1.9	7218	1	US-08-232-463-14	Sequence 14, Appl
C 2	61.2	1.6	51259	3	US-08-781-891-209	Sequence 209, App
C 3	58	1.5	16442	3	US-08-781-891-208	Sequence 208, App
4	54.8	1.4	1362	4	US-08-979-608A-12	Sequence 12, Appl
5	54.8	1.4	1422	4	US-08-979-608A-13	Sequence 13, Appl
6	54.8	1.4	1617	4	US-08-979-608A-11	Sequence 11, Appl
7	54.2	1.4	1926	4	US-09-249-585A-2	Sequence 2, Appli
8	54.2	1.4	2580	3	US-09-050-863-2	Sequence 2, Appli
9	54.2	1.4	2580	4	US-09-359-081-2	Sequence 2, Appli
C 10	54.2	1.4	5452	2	US-09-130-114-1	Sequence 1, Appli
11	54.2	1.4	9600	4	US-08-410-647-1	Sequence 1, Appli
12	54.2	1.4	9600	4	US-09-620-925-1	Sequence 1, Appli
13	54.2	1.4	10596	1	US-07-884-811-15	Sequence 15, Appl
14	54.2	1.4	10596	1	US-07-885-971-15	Sequence 15, Appl
15	54.2	1.4	10596	1	US-08-087-783A-15	Sequence 15, Appl
16	54.2	1.4	10596	1	US-08-194-088B-15	Sequence 15, Appl
17	54.2	1.4	10596	2	US-08-194-087-15	Sequence 15, Appl
18	54.2	1.4	10596	5	PCT-US93-04648-15	Sequence 15, Appl
19	53.4	1.4	1276	4	US-09-177-325-2	Sequence 2, Appli
20	53.4	1.4	1276	4	US-09-411-812A-2	Sequence 2, Appli
21	53.4	1.4	1276	4	US-09-590-113-2	Sequence 2, Appli
22	51.4	1.4	3489	2	US-08-728-323A-1	Sequence 1, Appli
23	51.4	1.4	3489	4	US-09-298-568-1	Sequence 1, Appli
C 24	51.4	1.4	32207	2	US-08-770-379-20	Sequence 20, Appl
C 25	51.4	1.4	32207	4	US-08-757-669A-20	Sequence 20, Appl
C 26	51.4	1.4	32207	4	US-09-230-371A-20	Sequence 20, Appl
27	51	1.3	2301	1	US-08-306-691B-23	Sequence 23, Appl
28	51	1.3	2301	5	PCT-US93-06251-78	Sequence 78, Appl
29	50.4	1.3	1678	3	US-08-650-766-2	Sequence 2, Appli
30	50.4	1.3	1954	4	US-08-922-635-2	Sequence 2, Appli
31	50.4	1.3	3318	3	US-08-650-766-3	Sequence 3, Appli
32	50.4	1.3	3318	3	US-08-922-635-3	Sequence 3, Appli
33	50.4	1.3	3385	3	US-08-650-766-1	Sequence 1, Appli
34	50.4	1.3	3385	3	US-08-922-635-1	Sequence 1, Appli
35	50.4	1.3	15202	3	US-08-922-635-21	Sequence 21, Appl
36	49	1.3	19056	4	US-09-272-032-8	Sequence 8, Appli
37	48.8	1.3	3853	4	US-09-245-248B-53	Sequence 53, Appl
38	47.4	1.3	2313	4	US-09-370-838-157	Sequence 157, App
39	47.4	1.3	9551	1	US-08-056-200-93	Sequence 93, Appl
40	47.4	1.3	9551	2	US-08-800-644-93	Sequence 93, Appl
C 41	47.2	1.2	1314	4	US-09-500-569-5	Sequence 5, Appli
C 42	47	1.2	58	4	US-08-860-038-16	Sequence 16, Appl
C 43	47	1.2	58	4	US-09-580-923-16	Sequence 16, Appl
44	46.8	1.2	58	4	US-08-860-038-15	Sequence 15, Appl
45	46.8	1.2	58	4	US-09-580-923-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 1.9%; Score 73.4; DB 1; Length 7218;

Best Local Similarity 5.0%; Pred. No. 2e-06;
Matches 27; Conservative 225; Mismatches 146; Indels 0; Gaps 0;

Qy 145 GAAGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGCTTCCAGCAGAAACGGAACTATTTC 204
Db 1429 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1370

Qy 205 GAGAGATCCCCACCCA-AGATCACAGTGAAGAGAATGAAGCCCCCGCCCAAGGTCCAC 264
Db 1369 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1310

Qy 265 ATCTCCCAGGGGAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGCAG 324
Db 1309 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1250

Qy 325 CACAAGAAGTGGAAACAAGAGCAGCGCCTCCCTCCTCCCATCGTGAAGAAG 384
Db 1249 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1190

Qy 385 TCA-CGAAGAAGGAGGTCAAAGCTCCAAAAGAGCCTGAAGCTGAGCCAAGAG 444
Db 1189 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1130

Qy 445 AAAGTGCCCAAGGCTCATGGCCAAGGGCCAGCCCAAGGCAGGGGACTCTGTGTCCCACAT 504
Db 1129 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1070

Qy 505 CAGATTCCAAGCCCAAGCGGCCACAACCCAGCAGGGAAATT 545
Db 1069 RRRATTCCAAGCTCCCTCGACTTGACCAAGCTCCGAAATT 1029

RESULT 2
US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No, 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 24-852.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51759 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-209

Query Match 1.6%; Score 61.2; DB 4; Length 51759;
Best Local Similarity 51.9%; Pred. No. 0.0014;
Matches 138; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1 AAGTTTTTTTACAATTTTTTGAGAAGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60
Db 283 AAACAATAACAAATAATACATGAAGAAATGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 224

Qy 61 AGGGAGAACTTCCTAAGGAGAGTCGAAGACAAGCCTAACAAGAAAAAGAAGAAGAAG 120
Db 223 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 164

Qy 121 GGAAATAATTTCTAGAAGAAGGGGGAAGCCAGAGATCAGAGTGAAGTGAATGAAGGGGGG 180
Db 163 GAAGAAGAAGAAGAAGGAGGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 104

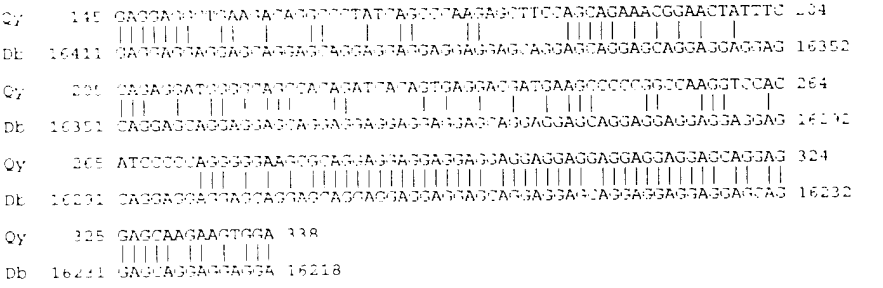
Qy 181 GGAAAGTTCACAACCAAGAAGAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 240
Db 103 GAAGGAGGAAGGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG 44

Qy 241 GAGGAGGAGGAGGCAAGAAGTGGA 338
Db 43 GAGGAGGAGGAGGAGGCAGGAGGA 18

RESULT 3
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No, 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 24-852.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-208

Query Match 1.5%; Score 58; DB 3; Length 16442;

Best Local Similarity 56.2%; Pred. No. 0.0047;
Matches 109; Conservative 0; Mismatches 85; Indels 0; Gaps 0;



RESULT 4
US-08-979-608A-12
; Sequence 12, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-No 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 12
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1362 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...696
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-08-979-608A-12

Query Match 1.4%; Score 54.8; DB 4; Length 1362;

Best Local Similarity 48.1%; Pred. No. 0.01;
Matches 155; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

RESULT 5
US-08-979-608A-13
; Sequence 13, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-No. 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1422 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..756
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-979-608A-13

Query Match 1.4%; Score 54.8; DB 4; Length 1422;
Best Local Similarity 48.1%; Pred. No. 0.01;
Matches 155; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 277 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGTG 336
Db 112 GAAAAGGAGGAGGAGGAGGAGGAGGAGGAAGACGAGGAGGAGGAGGAGGAGGAGGTGGTG 171

Qy 337 GAAACAAGAGCAGCGGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCATTGAAGCAA 396
Db 172 TCCGAGGGCTCGGAGGTGCCCGAGAGCGATCGTCCCCGCGGTGCGCAGCATCACCAGCTG 231

Qy 397 GGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAG 456
Db 232 AATGGCGGCGAGCGCGGCCCGCAGACCGCCAAGGAGCGGGCCAAGGAGTGGTCGCTGTGT 291

Qy 457 GGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTGCGCAGTCAGACTCCAAGC 516
Db 292 GACCCTGGCCAGGAGGAAGGCGGGGGCCGCCCGCGGGCAGTGGCACCCGCCAG 351

Qy 517 CCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGCATGTACCAGAGCCGCC 576
Db 352 GTGTTCTCCATGGCGGCCTTGAGTAAGGAGGGGGATCAGCCTCTTCGACCACCCGGCCT 411

Qy 577 CGGGCCCCGTGCCTGTGCCCGT 598
Db 412 GACTCCCGTCCCCGGTGCCTT 433

RESULT 9
US-08-979-608A-11
; Sequence 11, Application US/08979608A
; Patent No. 6358461
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-Nov-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/049,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,954
; REFERENCE/DOCKET NUMBER: 10759-000001 (formerly 3983/D3818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542 5070
; TELEFAX: 617/542 8906
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-08-979-608A-11

Query Match 1.4%; Score 54.8; DB 4; Length 1617;
Best Local Similarity 48.1%; Pred. No. 0.011;
Matches 155; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 277 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGTG 336
Db 307 GAAAAGGAGGAGGAGGAGGAGGAGGAGGAAGACGAGGAGGAGGAGGAGGAGGAGGTGGTG 366

Qy 337 GAAACAAGAGCAGCGGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCATTGAAGCAA 396
Db 367 TCCGAGGGCTCGGAGGTGCCCGAGAGCGATCGTCCCCGCGGTGCGCAGCATCACCAGCTG 426

Qy 397 GGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACAGCGGCCAAG 456
Db 427 AATGGCGGCGAGCGCGGCCCGCAGACCGCCAAGGAGCGGGCCAAGGAGTGGTCGCTGTGT 486

Qy 457 GGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTGCGCAGTCAGACTCCAAGC 516
Db 487 GGCCCTCACCCTGGCCAGGAGGAAGGCGGGGGGCCGCCCGCGGGCAGTGGCACCCGCCAG 546

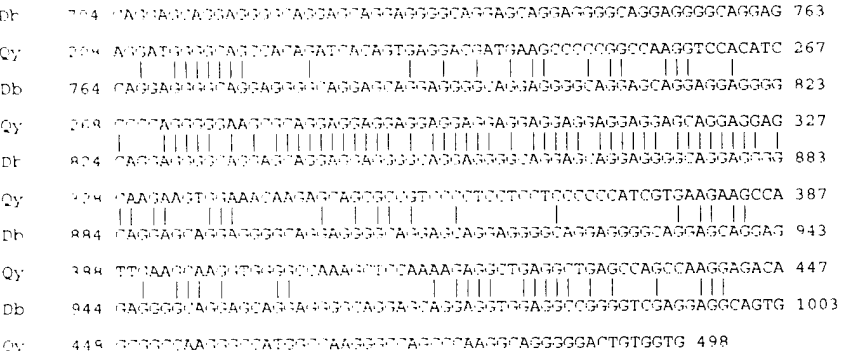
Qy 517 CCAAGCGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGCATGTACCAGAGCCGCC 576
Db 547 GTGTTCTCCATGGCGGCCTTGAGTAAGGAGGGGGGATCAGCCTCTTCGACCACCCGGCCT 606

Qy 577 CGGGCCCCGTGCCTGTGCCCGT 598
Db 607 GACTCCCGTCCCCGGTGCCTT 628

RESULT 7
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Burlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND [illegible]
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: Coding strand of EBNA 1 DNA
US-09-249-585A-2

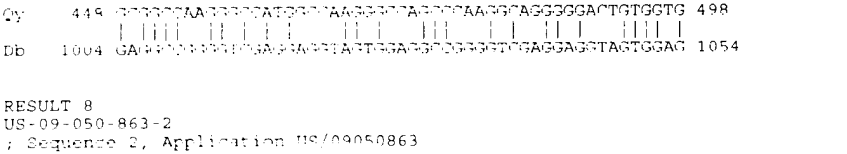
Query Match 1.4%; Score 54.2; DB 4; Length 1926;
Best Local Similarity 45.7%; Pred. No. 0.015;
Matches 180; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

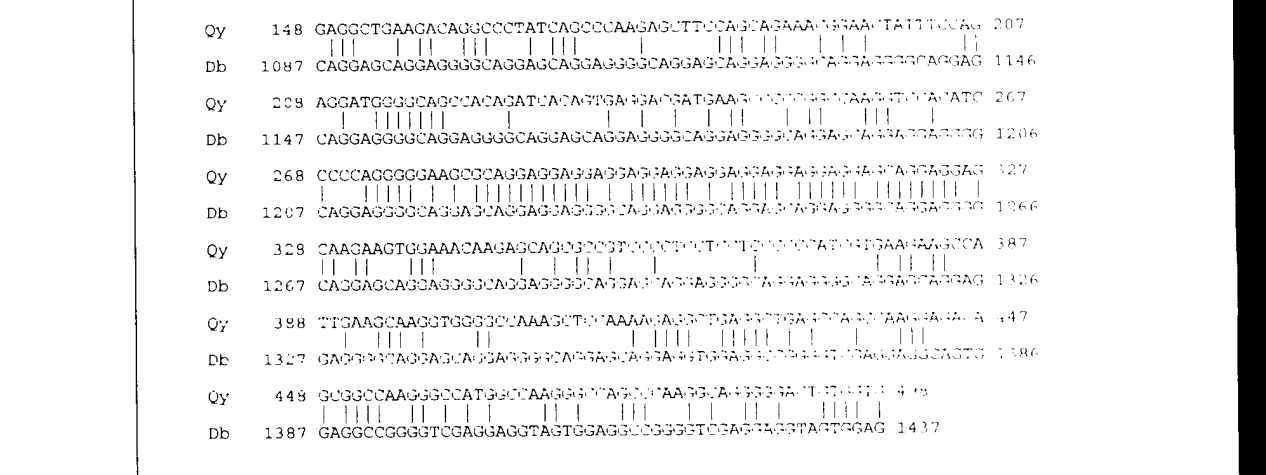
Qy 148 GAGGCTCAAGACAGGCCTATAGCTAAGAGCTTTATCTATTTTTAA 197
Db 644 GAGGAGGGGCAGGAGCAGGAGGAGGAGGCGAGGAGGCAGAGGCAGAGGAGGG 703



RESULT 8
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; TITLE OF INVENTION: System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-09-050-863-2

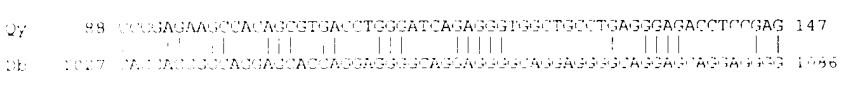
Query Match 1.4%; Score 54.2, DB 3, Length 2580;
Best Local Similarity 45.7%; Pred No 0.016;
Matches 198; Conservative 6, Mismatches 223; Indels 0; Gaps 0;

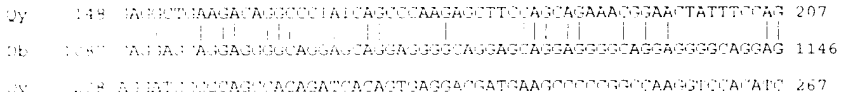


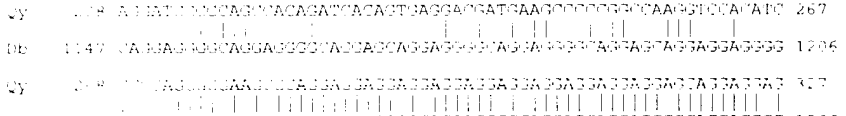


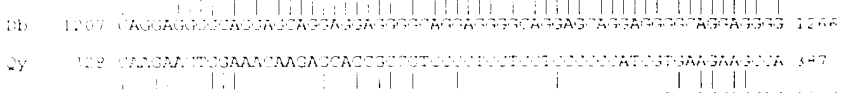
RESULT 9
US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; Hiang, Betty
; Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/359,081
; FILING DATE: 22-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/050,863
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

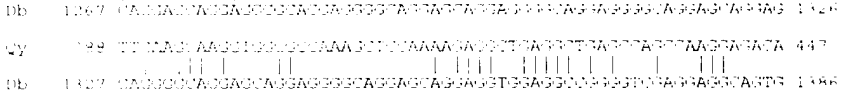
Query Match 1.4%; Score 54.2; DB 4; Length 2580;
Best Local Similarity 45.7%; Pred. No. 0.016;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

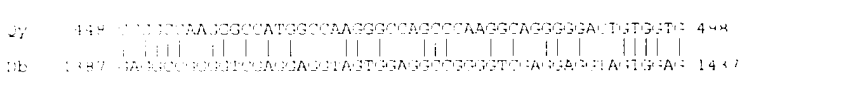






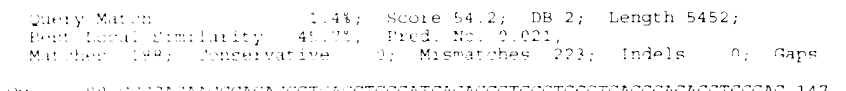


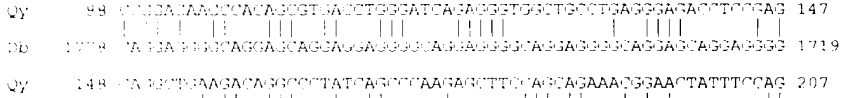


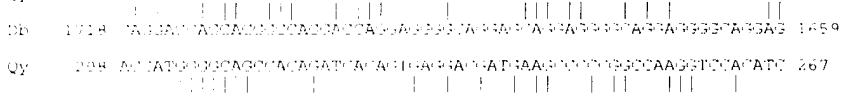


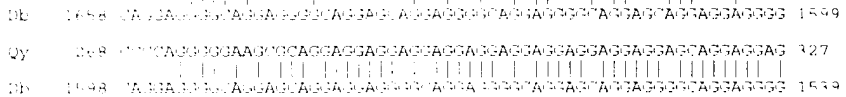
RESULT 10
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Berlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: from Multiple Transfected Episomes
; FILE REFERENCE: 9869/1900[illegible]
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 2.0
; SEQ ID NO 2
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEENA
US-09-130-114-2

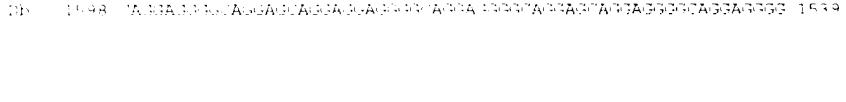
Query Match 1.4%; Score 54.2; DB 2; Length 5452;
Best Local Similarity 45.7%; Pred. No. 0.021;
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

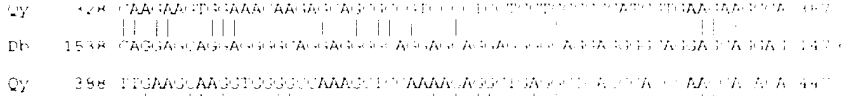


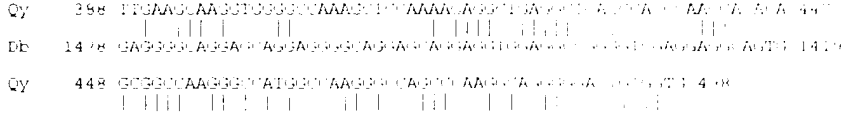


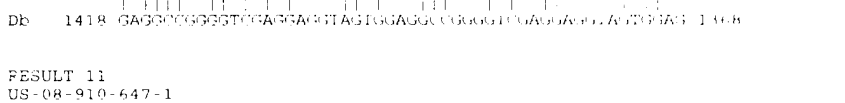






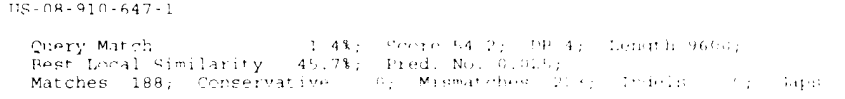


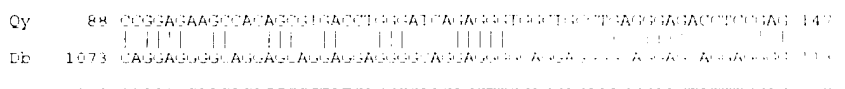


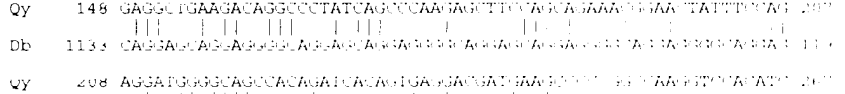


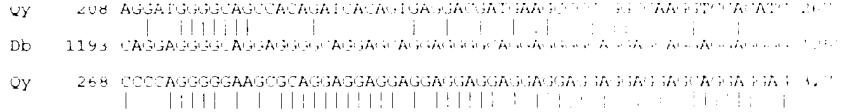
RESULT 11
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2366
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match 1.4%; Score 54.2; DB 4; Length 9600;
Best Local Similarity 45.7%; Pred. No. 0.03[illegible];
Matches 188; Conservative 0; Mismatches 223; Indels 0; Gaps 0;









```
Db    1253 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGG 1312

Qy     328 CAAGAAGTGGAAACAAGAGCAGCGGGGTCCCCTCCTCCTCCCCCCATCGTGAAGAAGCCA 387
           || ||   |||         |  | || |   |         |       | || ||
Db    1313 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 1372

Qy     388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447
             | ||| |   ||          | ||||  ||||| | |   |  |||
Db    1373 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGGTCGAGGAGGCAGTG 1432

Qy     448 GCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTG 498
           | ||||  || | | |    || |  |||   | |  || |   ||||
Db    1433 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 1483


RESULT 12
US-09-620-925-1
; Sequence 1, Application US/09620925
; Patent No. 6468986
;  GENERAL INFORMATION:
;       APPLICANT: Zuckermann et al.
;       TITLE OF INVENTION: Compositions and Methods for
;                           Polynucleotide Delivery
;       NUMBER OF SEQUENCES: 4
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Chiron Corporation
;            STREET: 4560 Horton Street
;            CITY: Emeryville
;            STATE: California
;            COUNTRY: U.S.A.
;            ZIP: 94608-2916
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/620,925
;            FILING DATE: 21-Jul-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/910,647
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Fujita, Sharon M.
;            REGISTRATION NUMBER: 38,459
;            REFERENCE/DOCKET NUMBER: 1218.002
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (510) 923-2726
;            TELEFAX: (510) 655-3542
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 9600 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

  Query Match             1.4%;  Score 54.2;  DB 4;  Length 9600;
  Best Local Similarity   45.7%;  Pred. No. 0.025;
  Matches  188;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy      88 CCGGAGAAGCCACAGCGTGACCTGGGATCAGAGGGTGGCTGCCTGAGGGAGACCTCCGAG 147
           | ||||  ||   ||| ||   |||   |||||       | ||||          ||
Db    1075 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGG 1134

Qy     148 GAGGCTGAAGACAGGCCCTATCAGCCCAAGAGCTTCCAGCAGAAACGGAACTATTTCCAG 207
           |||   | ||  |||   | |||     |    ||| ||   | | |         ||
Db    1135 CAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAG 1194

Qy     208 AGGAIGGGGCAGCCACAGAGCACAGTGAGGAGGATGAAGCCCCGGCCAAGGTCCACATC 267
```

```
Db    1193 CAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 1252

Qy     268 CCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCAGGAGGAG 327
           |  ||||| | | ||||||||||| | |||||| | ||||| |||||| |||||||| |
Db    1253 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGG 1312

Qy     328 CAAGAAGTGGAAACAAGAGCAGCGGGGTCCCCTCCTCCTCCCCCCATCGTGAAGAAGCCA 387
           || ||   |||         |  | || |   |         |       | || ||
Db    1313 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGCAGGAG 1372

Qy     388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447
             | ||| |   ||          | ||||  ||||| | |   |  |||
Db    1373 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGGTCGAGGAGGCAGTG 1432

Qy     448 GCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTG 498
           | ||||  || | | |    || |  |||   | |  || |   ||||
Db    1433 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 1483


RESULT 13
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/884,811
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-884-811-15

  Query Match             1.4%;  Score 54.2;  DB 1;  Length 10596;
  Best Local Similarity   45.7%;  Pred. No. 0.026;
  Matches  188;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy      88 CCGGAGAAGCCACAGCGTGACCTGGGATCAGAGGGTGGCTGCCTGAGGGAGACCTCCGAG 147
           | ||||  ||   ||| ||   |||   |||||       | ||||          ||
Db    2607 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGG 2666

Qy     148 GAGGCTGAAGACAGGCCCTATCAGCCCAAGAGCTTCCAGCAGAAACGGAACTATTTCCAG 207
           |||   | ||  |||   | |||     |    ||| ||   | | |         ||
```

```
Db    2667 CAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAG 2726

Qy     209 AGGATGGGGCAGCCACAGATCACAGTGAGGACGATGAAGCCCCGGGCCAAGGTCCACATC 267

Db    2727 CAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2786

Qy     268 CCCCAGGGGGAAGGGCAGGAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGCAGGAGGAG 327

Db    2787 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGAG 2846

Qy     328 CAAGAAGTGGAAACAAGAGCAGCGGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCA 387

Db    2847 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2906

Qy     388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447

Db    2907 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGGTCGAGGAGGCAGTG 2966

Qy     448 GGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTG 498

Db    2967 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 3017


RESULT 14
US-07-885-971-15
; Sequence 15, Application US/07885971
; Patent No. 5328837
;  GENERAL INFORMATION:
;    APPLICANT:  Gilbert, Paul J.;  Lokker, Nathalie A.;  Mark, Melanie F.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR ISOFORM VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,971
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Draper, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3016
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-885-971-15

  Query Match             1.4%;  Score 54.2;  DB 1;  Length 10596;
  Best Local Similarity   45.7%;  Pred. No. 9.02e-06;
  Matches  198;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy      88 CCGGAGAAGCCACAGCGTGACCTGGAATCAGAGGGTGGCTGCCTGAGGGAGACCTCCGAG 147

Db    2607 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCAGGAGCAGGAGGAG 2666

Qy     148 GAGGCTGAAGACAGGCCCATCAGCCAAGACGTTCCAGCAGAAACCGAACTATTCCAG 207

Db    2667 CAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2726

Qy     208 AGGATGGGGCAGCCACAGATCACAGTGAGGACGATGAAG                    247

Db    2727 CAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2786

Qy     268 CCCCAGGGGGAAGGGCAGGAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGCAGGAGGAG 327

Db    2787 CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGAG 2846

Qy     328 CAAGAAGTGGAAACAAGAGCAGCGGCCGTCCCCTCCTCCTCCCCCATCGTGAAGAAGCCA 387

Db    2847 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2906

Qy     388 TTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAGGCTGAGGCTGAGCCAGCCAAGGAGACA 447

Db    2907 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGGTCGAGGAGGCAGTG 2966

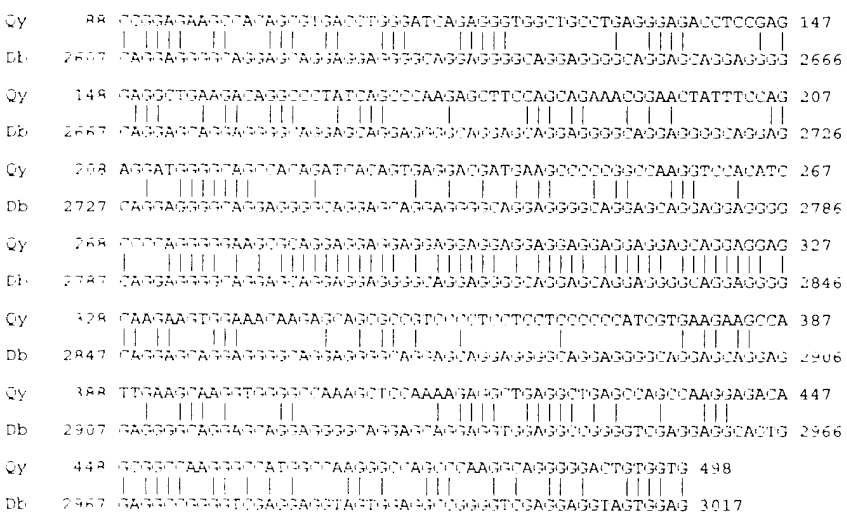
Qy     448 GGGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGGCAGGGGGACTGTGGTG 498

Db    2967 GAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAG 3017


RESULT 15
US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5547856
;  GENERAL INFORMATION:
;    APPLICANT:  Gilbert, Paul J., Lokker, Nathalie A., Mark, Melanie F.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/087,783A
;      FILING DATE:  13-Jul-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18 MAY-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marschang, Diane L.
;      REGISTRATION NUMBER:  36,600
;      REFERENCE/DOCKET NUMBER:  P0755779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5416
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-087-783A-15

  Query Match             1.4%;  Score 54.2;  DB 1;  Length 10596;
```

Best Local Similarity 45.7%; Pred. No. 0.026;
Matches 198, Conservative 0, Mismatches 223, Indels 0; Gaps 0,



Search completed: January 10, 2003, 02:42:12
Job time : 408.379 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:18:37 ; Search Time 145.541 Seconds
(without alignments)
11430.411 Million cell updates/sec

Title: US-09-803-126-7
Perfect score: 3780
Sequence: 1 cggcagcaggctcggg[illegible]aaaaaaaaaaaaaaaaaaa 3780

Scoring table: IDENTITY_NUC
Gap: 10.0 , GapExt 1.0

Searched: 380080 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3780	100.0	3780	10	US-09-803-126-7	Sequence 7, Appli
2	3352	88.7	4174	10	US-09-803-126-5	Sequence 5, Appli
3	1894.4	50.1	1960	10	US-09-822-849A-265	Sequence 265, App
4	1420	37.6	6293	10	US-09-803-126-2	Sequence 2, Appli
5	954.4	25.2	1247	9	US-09-764-868-145	Sequence 145, App
6	946.4	25.0	4375	10	US-09-803-126-3	Sequence 3, Appli
c 7	442	11.7	24757	9	US-09-764-868-1347	Sequence 1347, Ap
c 8	292	7.7	311	10	US-09-954-456-176	Sequence 176, App
c 9	292	7.7	311	10	US-09-954-456-861	Sequence 861, App
c 10	292	7.7	311	10	US-09-880-107-1520	Sequence 1520, Ap
c 11	174	4.6	526	10	US-09-864-761-7087	Sequence 7087, Ap
c 12	170	4.5	170	10	US-09-864-761-23818	Sequence 23818, A
c 13	137.8	3.6	233	10	US-09-864-761-24717	Sequence 24717, A
c 14	137.8	3.6	566	10	US-09-864-761-8013	Sequence 8013, Ap
15	113.8	3.0	312	10	US-09-783-590-1780	Sequence 1780, Ap
16	102.4	2.7	145	10	US-09-783-590-3077	Sequence 3077, Ap
17	59.2	1.6	3826	12	US-10-002-600-7	Sequence 7, Appli
18	57	1.5	3809	12	US-10-001-870-68	Sequence 68, Appl
c 19	56.4	1.5	650158	9	US-09-771-208-20	Sequence 20, Appl
c 20	55.8	1.5	234	10	US-09-864-761-19061	Sequence 19061, A
c 21	55.8	1.5	468	10	US-09-864-761-2326	Sequence 2326, Ap
c 22	55.4	1.5	267	10	US-09-864-761-27984	Sequence 27984, A
c 23	55.4	1.5	53226	10	US-09-818-264-3	Sequence 3, Appli
24	54.8	1.4	1362	9	US-09-976-740-12	Sequence 12, Appl
25	54.8	1.4	1362	10	US-09-962-055-12	Sequence 12, Appl
26	54.8	1.4	1362	12	US-10-023-529-12	Sequence 12, Appl
27	54.8	1.4	1362	12	US-10-023-523-12	Sequence 12, Appl
28	54.8	1.4	1422	9	US-09-976-740-13	Sequence 13, Appl
29	54.8	1.4	1422	10	US-09-962-055-13	Sequence 13, Appl
30	54.8	1.4	1422	12	US-10-023-529-13	Sequence 13, Appl
31	54.8	1.4	1422	12	US-10-023-523-13	Sequence 13, Appl
32	54.8	1.4	1617	9	US-09-976-740-11	Sequence 11, Appl
33	54.8	1.4	1617	10	US-09-962-055-11	Sequence 11, Appl
34	54.8	1.4	1617	12	US-10-023-529-11	Sequence 11, Appl
35	54.8	1.4	1617	12	US-10-023-523-11	Sequence 11, Appl
36	54.8	1.4	2561	9	US-09-976-740-48	Sequence 48, Appl
37	54.8	1.4	2561	12	US-10-023-529-48	Sequence 48, Appl
38	54.8	1.4	2561	12	US-10-023-523-48	Sequence 48, Appl
c 39	54.6	1.4	474	10	US-09-864-761-11284	Sequence 11284, A
c 40	54	1.4	465237	10	US-09-933-267A-1	Sequence 1, Appli
c 41	53.4	1.4	360	10	US-09-864-761-28274	Sequence 28274, A
c 42	53.4	1.4	450	10	US-09-864-761-11696	Sequence 11696, A
c 43	53.4	1.4	475	10	US-09-864-761-6203	Sequence 6203, Ap
c 44	53.4	1.4	511	10	US-09-864-761-22817	Sequence 22817, A
45	53.4	1.4	4744	10	US-09-880-107-3459	Sequence 3459, Ap

ALIGNMENTS

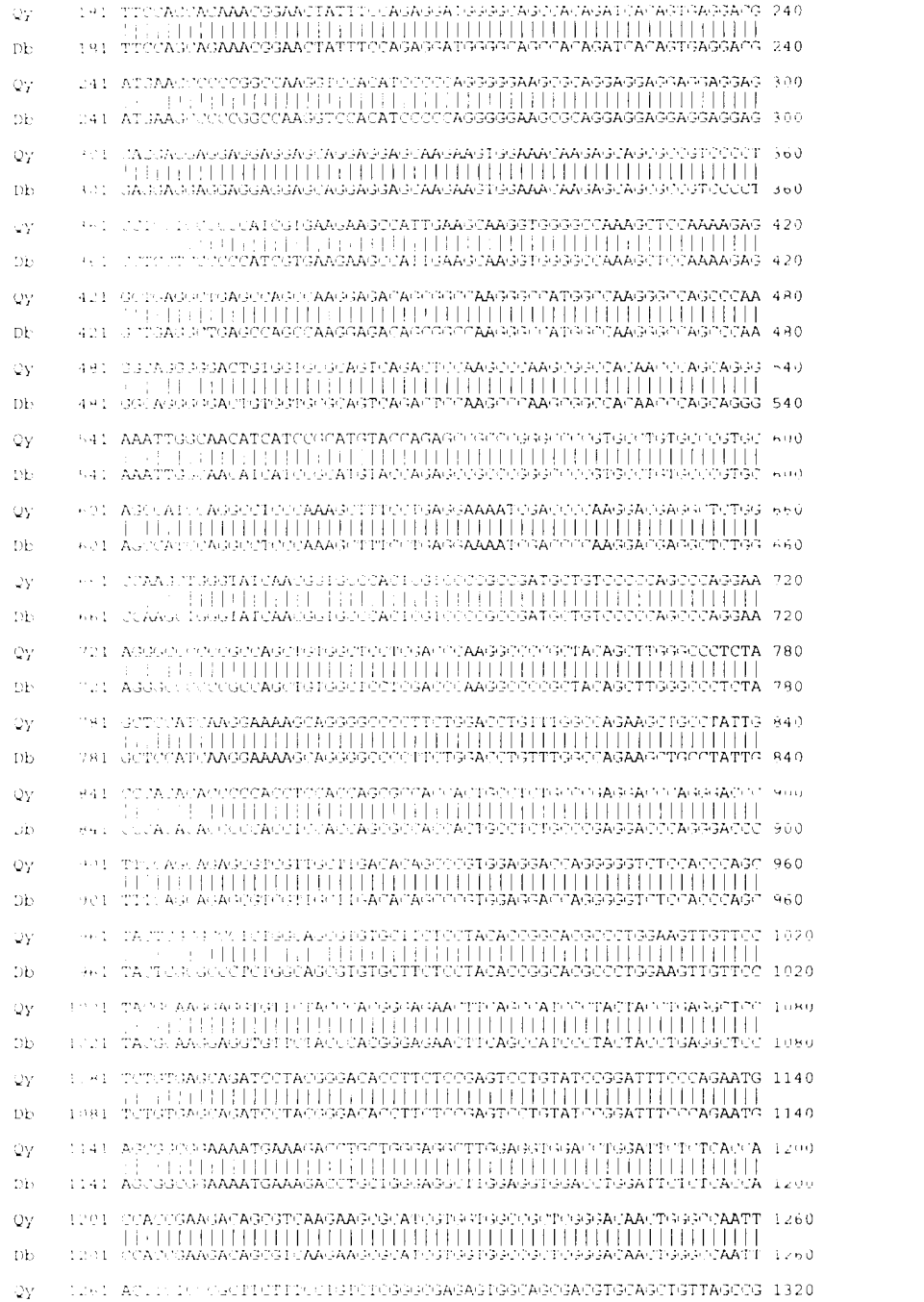
RESULT 1
US-09-803-126-7
; Sequence 7, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Pubanyi, Gabor M
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related
; OTHER INFORMATION: protein variant 2 (hMPP2)
US-09-803-126-7

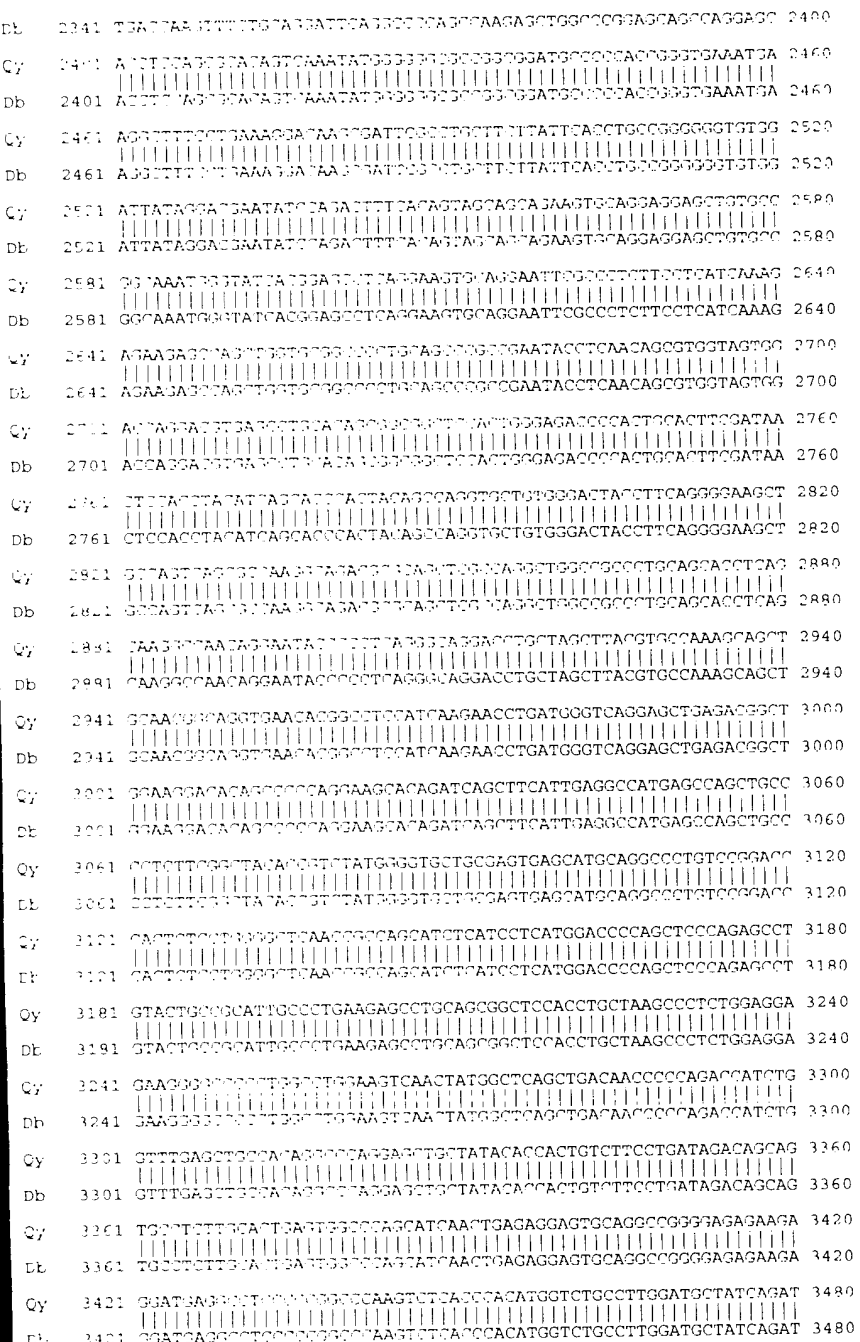
Query Match 100.0%; Score 3780; DB 10; Length 3780;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

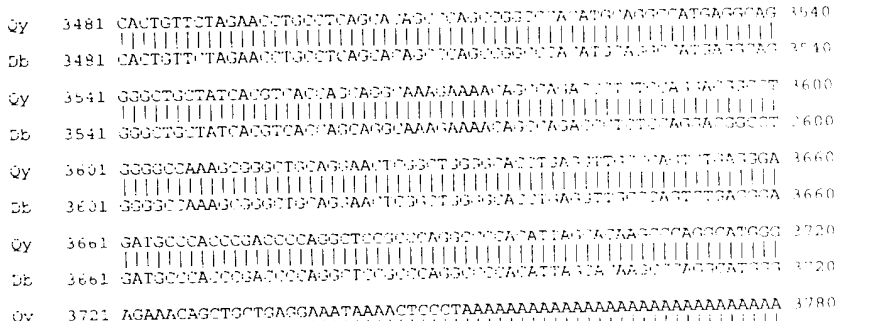
Qy 1 CGGCAGCAGCAGGCTCGGGCCTCCGAGGCTGCGTCCCAGGGCTCAGCCTCAGCCGTCACC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CGGCAGCAGCAGGCTCGGGCCTCCGAGGCTGCGTCCCAGGGCTCAGCCTCAGCCGTCACC 60

Qy 61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 TCCAAGCCCAGGAAGCCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAG 120

Qy 121 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GGTGGCTGCCTGAGGGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180

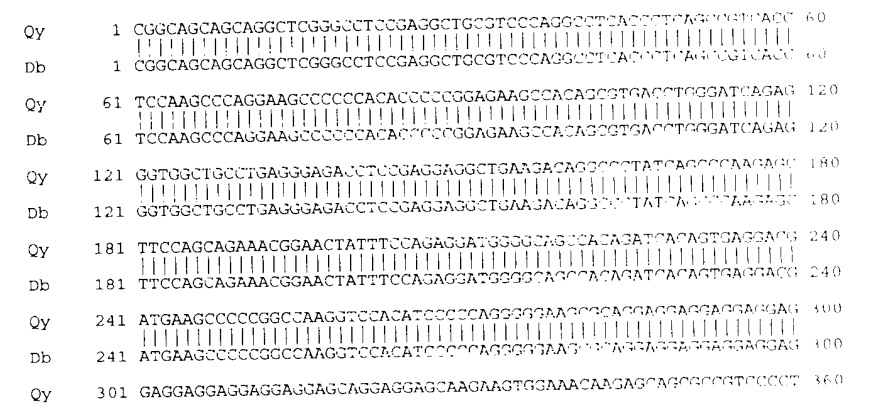


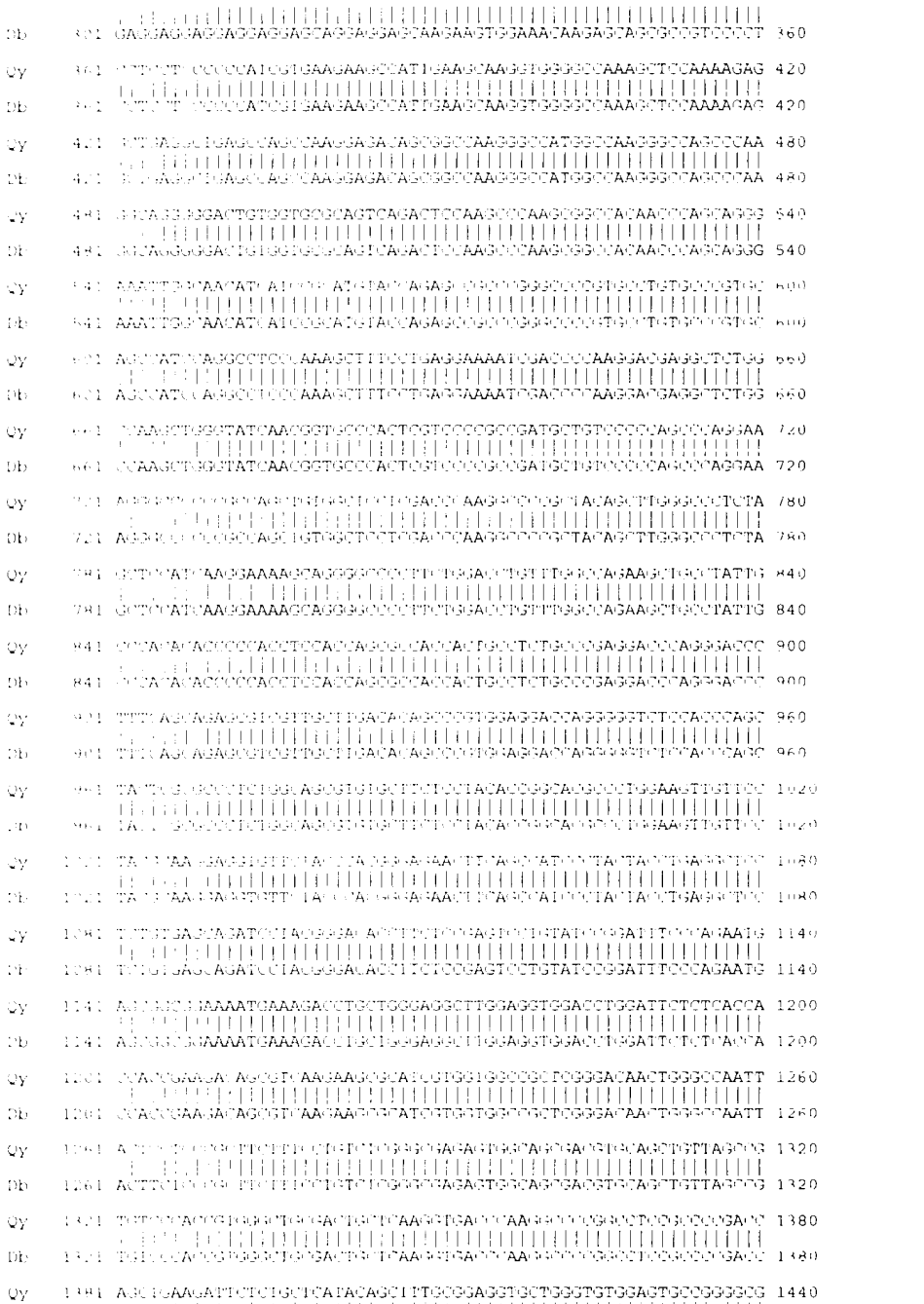




RESULT 2
US-09-803-126-5
; Sequence 5, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
, APPLICANT. Brooks, Alan R
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
, TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
, CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
, PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
, NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 5
; LENGTH: 4174
, TYPE. DNA
; ORGANISM: Homo sapiens
, FEATURE.
; OTHER INFORMATION: Partial DNA sequence for human myosin related
, OTHER INFORMATION: protein variant 1 (hMPP1)
US-09-803-126-5

Query Match 88.7%, Score 3352; DB 10; Length 4174;
Best Local Similarity 99.5%, Pred. No. 0.
Matches 3778; Conservative 0; Mismatches 0; Indels 396; Gaps 3;





```
Qy   2185 GGGTGACCAGTCCAAGCCCCGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAA 2185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GGGTGACCAGTCCAAGCCCCGGGCAAGGATGAGATGGATCTGCTCTATGAACTGCTGAA 2580

Qy   2186 GCTGTGCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGGTCAC 2245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GCTGTGCCAGCAGGAGAAGCTGAGGGATGAGATTTACTGCCAGGTTATCAAGCAGGTCAC 2640

Qy   2246 AGGACACCCCCGGGCCGGAACACTGCACTTGAGGCTGGAGCTTCCTCAGCCTTCTCACAGG 2305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AGGACACCCCCGGGCCGGAACACTGCACTTGAGGCTGGAGCTTCCTCAGCCTTCTCACAGG 2700

Qy   2306 CTTCTTCCCCCCGTCGACCAGGCTGATGCCCTACCTGACCAAGTTTCTGCAGGATTCAGG 2365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTTCTTCCCCCCGTCGACCAGGCTGATGCCCTACCTGACCAAGTTTCTGCAGGATTCAGG 2760

Qy   2366 CCCCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCACCTCCAGGGCACAGTCAAATATGG 2425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 CCCCAGCCAAGAGCTGGCCCGGAGCAGCCAGGAGCACCTCCAGGGCACAGTCAAATATGG 2820

Qy   2426 GGGGCGCCCGGGGGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGAT 2485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GGGGCGCCCGGGGGATGCCCCACCGGGTGAAATGAAGGCTTTCCTGAAAGGACAAGCGAT 2880

Qy   2486 TGGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATATCCAGACTTT 2545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TGGCCTGCTTCTTATTCACCTGCCGGGGGGTGTGGATTATAGGACGAATATCCAGACTTT 2940

Qy   2546 CACACTACAGCAGAAGTGCAGGAGGAGCTGTGCGGGCAAATGGGTATCACGGAGCCTCA 2605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CACAGTAGCAGCAGAAGTGCAGGAGGAGCTGTGCGGGCAAATGGGTATCACGGAGCCTCA 3000

Qy   2606 GGAAGTGCAGGAATTCGGCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGCGGCCCCT 2665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 GGAAGTGCAGGAATTCGGCCTCTTCCTCATCAAAGAGAAGAGCCAGCTGGTGCGGCCCCT 3060

Qy   2666 GCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCCG 2725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GCAGCCCGCCGAATACCTCAACAGCGTGGTAGTGGACCAGGACGTGAGCCTGCACAGCCG 3120

Qy   2726 GCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCAGCACCCACTA 2785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCTACATCAGCACCCACTA 3180

Qy   2786 CAGCCAGGTGCTGTGGGACTACGTTCAGGGGAAGCGTGCCAGTCAGCGCCAAGGCAGACGC 2845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 CAGCCAGGTGCTGTGGGACTACGTTCAGGGGAAGCGTGCCAGTCAGCGCCAAGGCAGACGC 3240

Qy   2846 GCAGCTCGGCCAGGCTGGCGGCCCTGCAGCAGCTCAGCAAGGCCAACAGGAATACCCCTC 2905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCAGCTCGGCCAGGCTGGCGGCCCTGCAGCAGCTCAGCAAGGCCAACAGGAATACCCCTC 3300

Qy   2906 AGGGCAGGACCTGGTAGCTTACGTGCCAAAGCAGCTGCAACGGCAGGTGAACACGGCCTC 2965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 AGGGCAGGACCTGGTAGCTTACGTGCCAAAGCAGCTGCAACGGCAGGTGAACACGGCCTC 3360

Qy   2966 CATCAAGAACCTGATGGGTCAGGAGCTGAGACGGCTGGAAGGACACAGCCCCCAGGAAGC 3025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CATCAAGAACCTGATGGGTCAGGAGCTGAGACGGCTGGAAGGACACAGCCCCCAGGAAGC 3420

Qy   3026 ACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGTCTATGG 3085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 ACAGATCAGCTTCATTGAGGCCATGAGCCAGCTGCCCCTCTTCGGCTACACCGTCTATGG 3480

Qy   3086 GGTGCTGCCAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGCCA 3145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 GGTGCTGCCAGTGAGCATGCAGGCCCTGTCCGGACCCACTCTCCTGGGGCTCAACCGCCA 3540

Qy   3146 GCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAGAG 3205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GCATCTCATCCTCATGGACCCCAGCTCCCAGAGCCTGTACTGCCGCATTGCCCTGAAGAG 3600
```

```
Qy   3206 CCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGGCCTGGAAGT 3265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CCTGCAGCGGCTCCACCTGCTAAGCCCTCTGGAGGAGAAGGGGCCCCCTGGCCTGGAAGT 3660

Qy   3266 CAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGTTGCCACAGGCCCAGGA 3325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 CAACTATGGCTCAGCTGACAACCCCCAGACCATCTGGTTTGAGTTGCCACAGGCCCAGGA 3720

Qy   3326 GCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACTGAGTGGCCCAG 3385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 GCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACTGAGTGGCCCAG 3780

Qy   3386 CATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCCTCCCCGGCCCAAG 3445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 CATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCCTCCCCGGCCCAAG 3840

Qy   3446 TCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGC 3505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 TCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGC 3900

Qy   3506 ACAGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGCTCACCAGCAG 3565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 ACAGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCAGCTCACCAGCAG 3960

Qy   3566 GCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGAAAC 3625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GCAAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGAAAC 4020

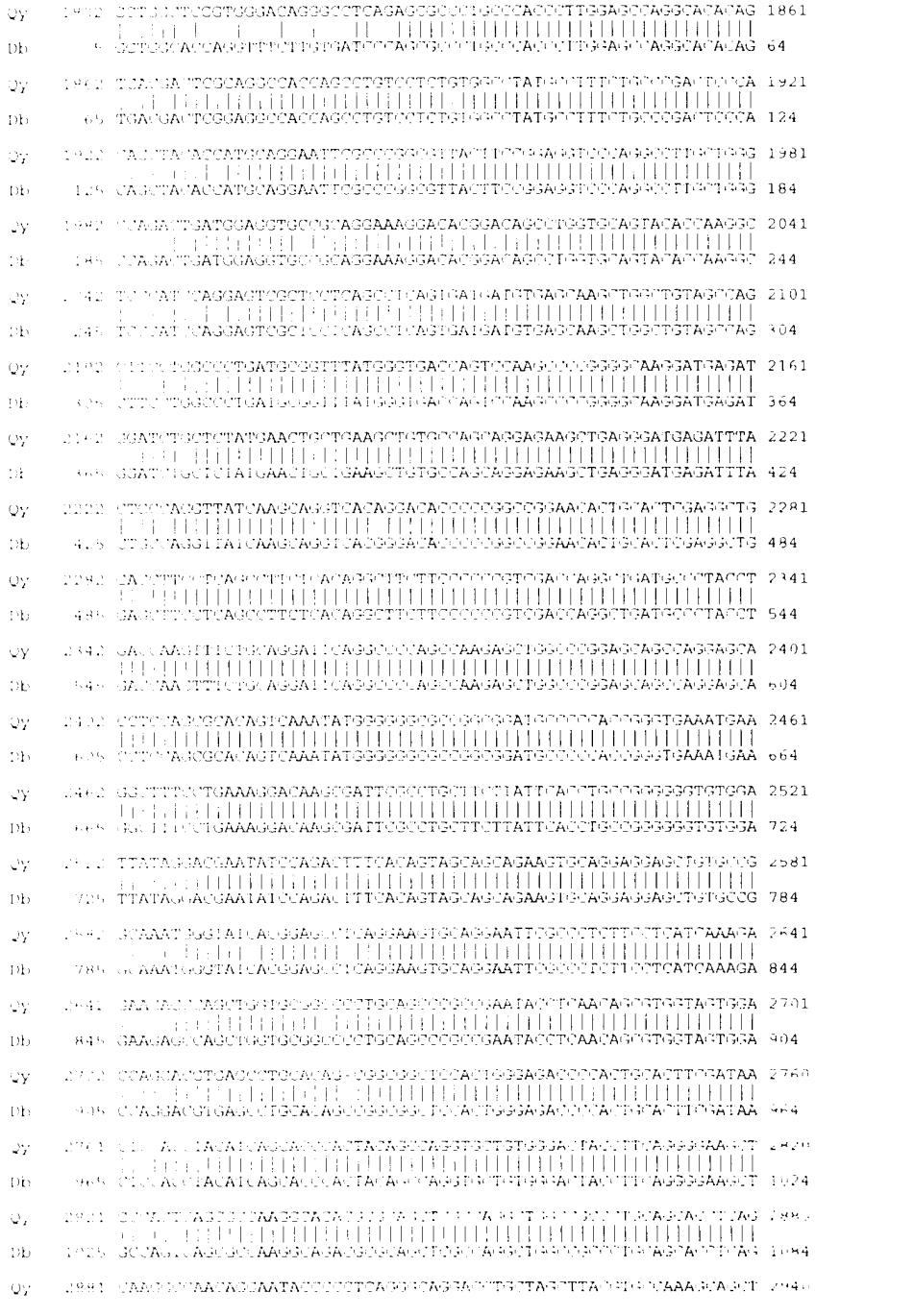
Qy   3626 TCGGCTGGGGCAGCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGAACCCAAGTCC 3685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 TCGGCTGGGGCAGCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCCGAACCCAAGTCC 4080

Qy   3686 GCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAA 3745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAA 4140

Qy   3746 ACTCCCTAAAAAAAAAAAAAAAAAAAAAAAAAAA 3778
          ||||||||||||||||||||||||||||||||||
Db   4141 ACTCCCTAAAAAAAAAAAAAAAAAAAAAAAAAAA 4174
```

RESULT 3
US-09-822-849A-265
; Sequence 265, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James P.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 265
; LENGTH: 1960
; TYPE: DNA
; ORGANISM: Homo sapiens
US 09-822-849A-265

Query Match 50.1%; Score 1894.4; DB 10; Length 1960;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1926; Conservative 0; Mismatches 16; Indels 2; Gaps 2;



RESULT 4
US-09-803-126-2
; Sequence 2, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Berg, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Polypept. Regulated Thr. Kinase, and Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015483-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09

```
;  PRIOR APPLICATION NUMBER: US 60/188,488
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;    LENGTH: 6293
;    TYPE: DNA
;    ORGANISM: Mus sp.
;    FEATURE:
;    OTHER INFORMATION: Mouse myosin related protein (MRP) variant 1 cDNA
US-09-803-126-2

  Query Match             39.4%;  Score 1490;  DB 10,  Length 6293;
  Best Local Similarity   67.1%;  Pred. No. 1.4e-296;
  Matches 2518;  Conservative    0;  Mismatches  810;  Indels  422;  Gaps   12,

Qy       4 CAGCAGCAGGCTCGGGCCTCGGAGGCTGCGTCCCAGGCCTCACCCTCAGCCGTCACCTCC 63
Db    2297 CAGCAGCAAGCTCAGACCTCTGGGGCCACCTCCCAGCCTCCACCCTCAACCACTGCTCCC 2356

Qy      64 AAGCCCAGGAAGCCCCCACACCCCCGGAGAAGCCACAGCGTGACCTGGGATCAGAGGGT 123
Db    2357 AAGCCCAAGAAGCCTCCTGTGCCCCCAAGAGAAGCCAGAGAGTAACCTAGAGCCTTCGGGT 2416

Qy     124 GGCTGCCTGAG---GGAGACCTCCGAGGAGGCTGAAGACAGGCCCTATCAGCCCAAGAGC 180
Db    2417 GTTGGCTTGAGAGAGGACACCCAGAGGAAGCTGAAAGCAAGCCTCAGCGCCCCAAGAGC 2476

Qy     181 TTCCAGCAGAAACGGAACTATTTCCAGAGGATGGGGCAGCCACAGATCACAGTGAGGACG 240
Db    2477 TTCCAACAGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACG 2536

Qy     241 ATGAAGCCCCCGGCCAAGGTCCACATCCCCCAGGGGGAAGCGCAGGAGGAGGAGGAGGAG 300
Db    2537 GTGAAACCTCCAGCCAAGGTTCAGATCCCCCAAGAGAGATGGAGGAGACGGAGGAGGAG 2596

Qy     301 GAGGAGGAGGAGGAGGAGGAGGAGCAAGAAGTGGAAACAAGAGCAGCGCCGTCCCCT 360
Db    2597 GAGGATGAGACCGCCGAGTTG----------  ---------TCCCCTCCTCCT 2629

Qy     361 CCTCCTCCCCCCATCGTGAAGAAGCCATTGAAGCAAGGTGGGGCCAAAGCTCCAAAAGAG 420
Db    2630 CCCCCTCCCCCGGTTGTGAAGAAGCCGCTGAAGGCAAGCAGGCCCAAAGCCGTAAAGGAA 2689

Qy     421 GCTGAGGCTGAGCCAGCCAAGGAG----------------------------------AC 446
Db    2690 GATGAGGCAGAGCCCGCCAAGAGGAAGTACCGACCCAGGGCGAGGATCCCCCGTGCAC 2749

Qy     447 AGCGGCCAAGGGCCATGGCCAAGGGCCAGCCCAAGCCAGGGGGACTGTGGTGCGCAGTCA 506
Db    2750 AGCTCCAACTCCGCACCTCAGCACCCCAAACCCAGCAGGTACCCCCAGTGCAGAGCTCC 2809

Qy     507 GACTCCAAGCCCAAGGGCCACAACCCAGCAGGGAAATTGGCAACATCATCCGCATGTAC 566
Db    2810 AACTCCGCAGTCTCAGCCCCCAACCCAGCAGGGAAATCCGAAACATCATCCGAATGTAC 2869

Qy     567 CAGAGCCCCCCGGGCCCCTGCCTGTGCCCGTGCAGCCATCCAGGCCTCCCAAAGCTTTC 626
Db    2870 CAGAGCCGTCCAGGCCTGTGGCTGTGCCCGTACAACCCACCAGGCCCATCAAAACTTTT 2929

Qy     627 CTGAGGAAAATCGACCCCAAGGACGAGGCTCTGGCCAAGCTGGGTATCAACGGTGCCCAC 686
Db    2930 CAGAAGAAAAATGACCCTAAGGATGAGGCTTTGGCTAAGTTAGGGATAAATGGCGTCCAC 2989

Qy     687 TCGTCCCCGCCGATGCTGTCCCCCAGCCCAGGAAAGGCCCCCCGCCAGCTGTGGCTCCT 746
Db    2990 TTGCCCCTATCCA---CATCGCCTAACCAAGGGAAGAGCTCTCCACCGGCTGTAGTTCCT 3046

Qy     747 CGACCCAAGGCCCCGTACAGCTTGGGCCTCTAGCTCCATCAAGGAAAAGCAGGGGCCC 806
Db    3047 CGACCTAAGGCTCGACCTGGTCTTGAGCCTTCCCTATCCATCCAGGAAAAGCAGGGACCC 3106

Qy     807 CTTCTGGACCTGTTTGGC----------CAGAAGCTGCCTATTGCCCACACACCCCCACCT 857

Db    3107 CTTCGGGACTTGTTTGGCCATGTAGTCCAAACCCACCTACAGCTCCAGCACCCCCGCCT 3166

Qy     858 CCACCAGCGCCACCACTGCCTCTGCCGGAGGACCCAGGGACCCTTTCAGCAGAGCGTCGT 917
Db    3167 CCACCAGCACTCCCACCGCCTCTGTCTGGGGAGCCCAAGACCCCTTCAGTGGAGTCCCAT 3226

Qy     918 TGCTTGACACAGCCCGTGGAGGACCAGGGGGTCTCCACCCAGCTACTCGGCGCCCTCTGGC 977
Db    3227 GCCTTGACAGAGCCCATGGAGGACAAGAACATCTCCACAAAGCCTCCTCTGCCCCTCTGGA 3286

Qy     978 AGCGTGTGCTTCTCCTACACCGGCACCGCCCTGGAAGTTGTTCCTACGCAAGGAGGTGTTC 1037
Db    3287 AGTGTGTGCTTCTCCTATGCCAATGCACCCTGGAAGTTGTTCTTACGCAAGGAGGTGTTC 3346

Qy    1038 TACCCACGGGAGAACTTCAGCCATCCCTACTACCTGAGGCTCCTCTGTGAGCAGATCCTA 1097
Db    3347 TACCCCCGGGAGAACTTCAGTCATCCATACTGCCTCAGTCTCCTCTGCCAGCAGATCCTG 3406

Qy    1098 CGGGACACCTTCTCCGAGTCCTGTATCCGGATTTCCCAGAATGAGCGGCGGAAAATGAAA 1157
Db    3407 CGGGACACCTTCACAGAGTCCTGCACCCGGATCTCACAGGATGAGCGGCACAAAATGAAA 3466

Qy    1158 GACCTGCTGGGAGGCTTGGAGGTGGACCTGGATTCTCTCACCACCACCGAAGACAGCGTC 1217
Db    3467 GGCCTTCTGGGAGACTTGGAGGTGAGTCTGGAGACCCTTGACATTGTTGAAGACAGCATC 3526

Qy    1218 AAGAAGCGCATCGTGGTGGCCGCTCGGGACAACTGGGCCAATTACTTCTCCCCCATCTTT 1277
Db    3527 AAAAAACGCATCGTGGTCGCTGCTCGGGACAACTGGGCCAATTACTTCTCCCCCATCTTC 3586

Qy    1278 CCTGTCTCGGGCGAGAGTGGCAGCGACGTGCAGCTGTTAGCCGTGTCCACCGTGGGCTG 1337
Db    3587 CCAGTCTCGGGTGAGAGTGGCAGCGATGTACAGCTGCTGGGTGTGTCTCACCGGGACTG 3646

Qy    1338 CGACTGCTCAAGGTGACCCAAGGCCCCGGCCTCCCCCCGACCAGCTGAAGATCTCTGC 1397
Db    3647 CGGCTGCTGAAGGTGACCCAAAGCCCGAGCTTCCACCTGGACCAGCTGAAGACACTCTGT 3706

Qy    1398 TCATACAGCTTTGCGGAGGTGCTGGGTGTGGAGTGCCGGGGGCGGCTCCACCCTGGAGCTG 1457
Db    3707 TCCTACAGCTATGCTGAAGTCCTGACCGTGCAGTGCAGGGGCAGATCCACCCTGGAGCTG 3766

Qy    1458 TCACTGAAGAGCGAGCAGCTGGTGCTGCACACAGCCCGGGCAAGGGCCATCGAGGCGCTG 1517
Db    3767 TCCTTGAAGAATGAGCAGCTGATACTGCACACAGCCTGGGCGAGGGCCATCAAGGCATG 3826

Qy    1518 GTTGAGCTATTCCTGAATGAGCTTAAGAAGGACTCCGGCTATGTCATCCCCCTGCGCAGC 1577
Db    3827 GTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTCATCGCCCTGCGCAGC 3886

Qy    1578 TACATCACTGACAACTGCAGCCTCCTCAGCTTCCACCGTGGGGACCTCATCAAGCTGCTG 1637
Db    3887 TACATCACCGATGACAATAGCCTCCTCAGTTTCCACCGTGGGGACCTCATTAGGTTACTG 3946

Qy    1638 CCGGTGGCCACCCTGGAGCCAGGCTGGCAGTTTGGCTCTGCCGGGGGCCGTTCCGGACTC 1697
Db    3947 CCAGTGACCGCTCTGGAACCAGGCTGGCAGTTCGGTTCTGCCGGGGGCCGTTCCGGACTC 4006

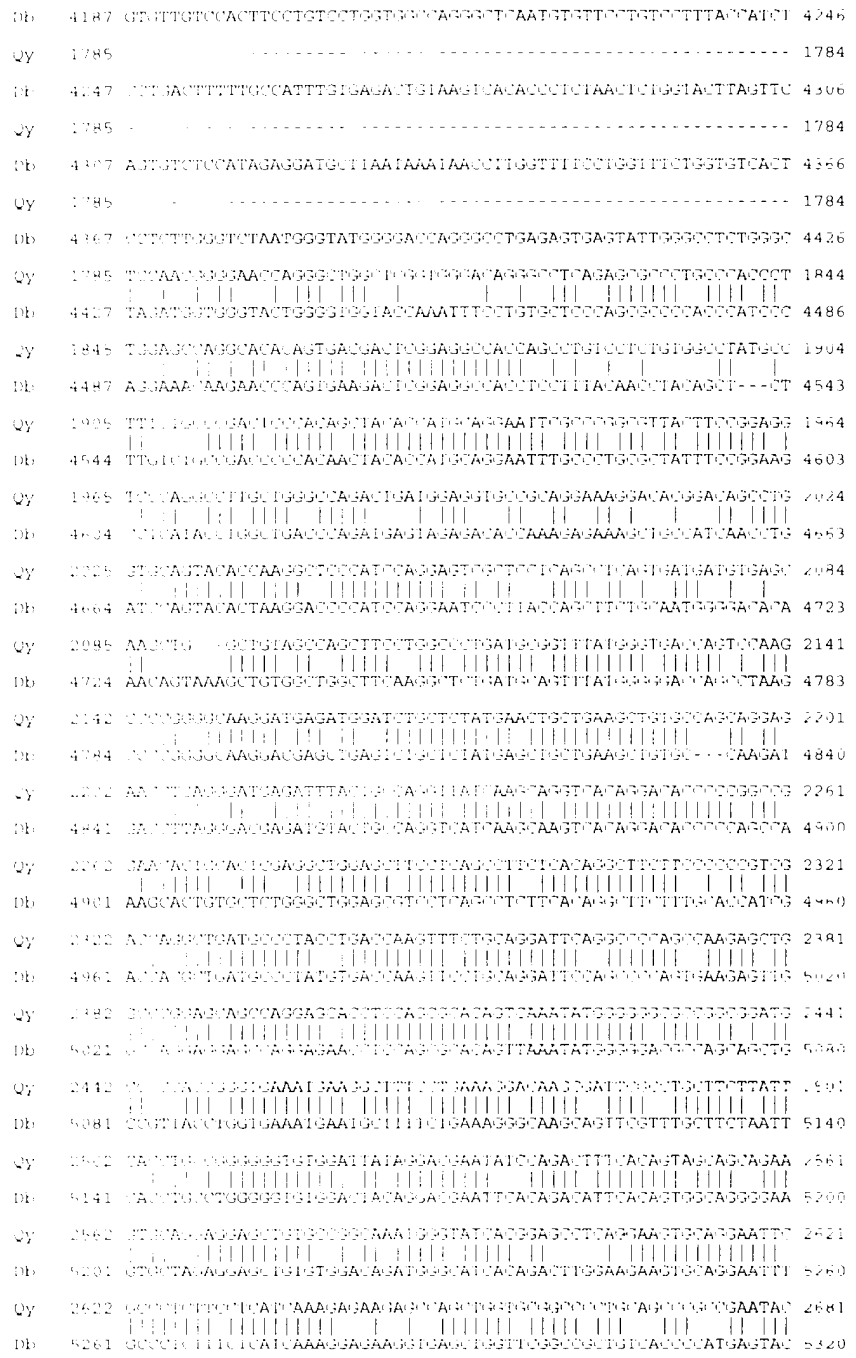
Qy    1698 TTTCCTGCCGACATAGTGCAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAG 1757
Db    4007 TTTCCCGATGACGTGGTGCAGCCAGCTGCTGCCCCCGACCCTCCTTTTCCCTGGGAAAG 4066

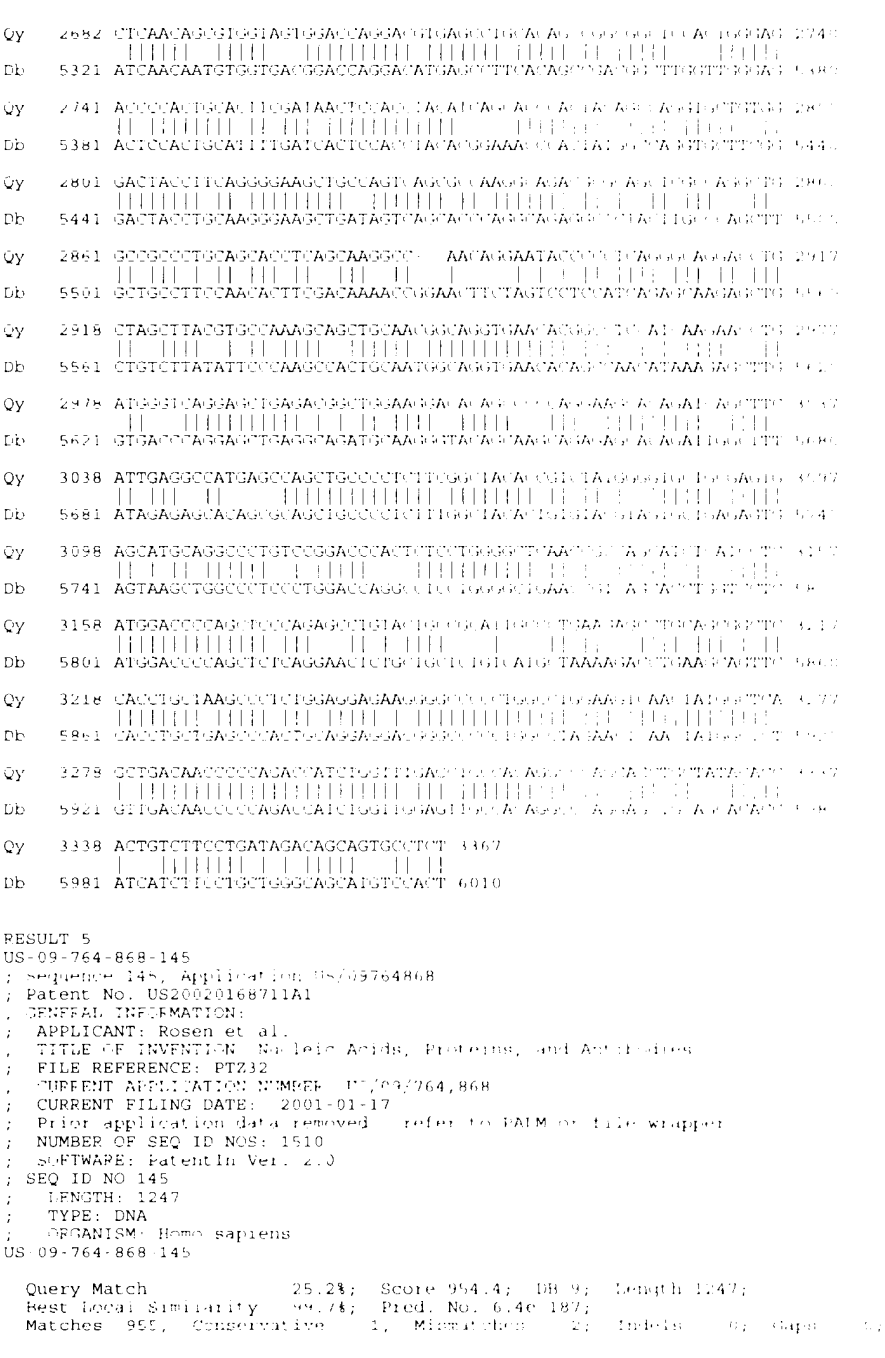
Qy    1758 AGGAGTGGCTGGCA---CAAGGGTCAGCTG---------------------------- 1784
Db    4067 AGAAACAGCTGGCAACGCAAGAGTAAGCTGCGGCCAGCTCAGGAGGTGAGGAAGACAGAA 4126

Qy    1785 ------------------------------------------------------------ 1784

Db    4127 GAGGTGAAGTGATACAGGCCTAACTTGGAGACTGAGAAGGAAAGAGCAGGGTTGCTTCGG 4186

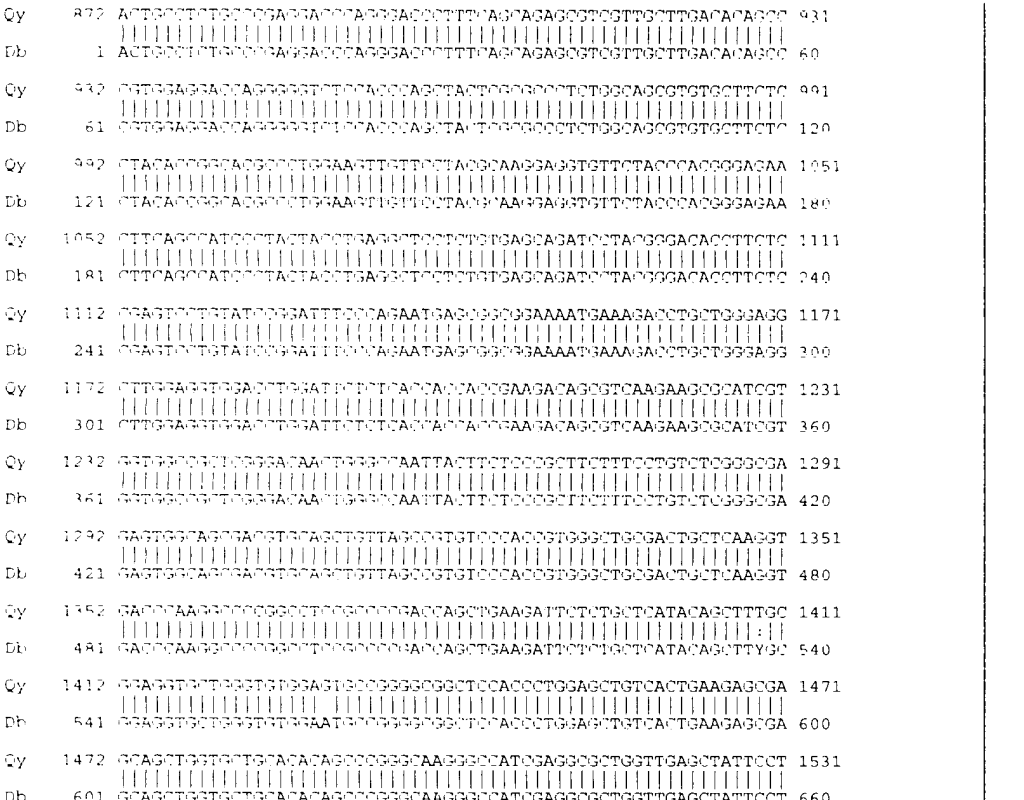
Qy    1785 ------------------------------------------------------------ 1784
```





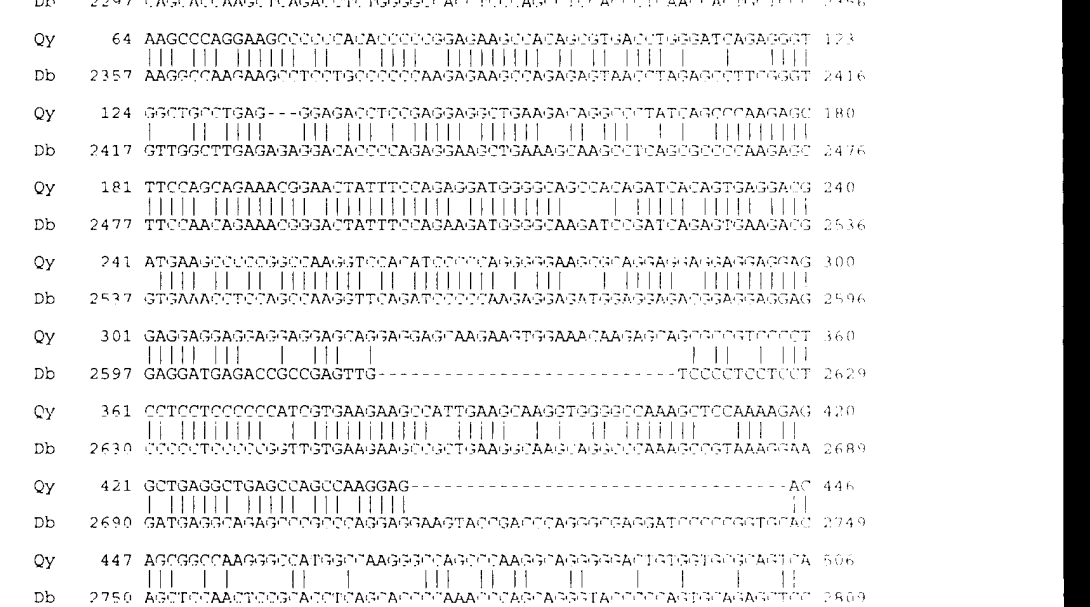
RESULT 5
US-09-764-868-145
; Sequence 145, Application: US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 145
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-145

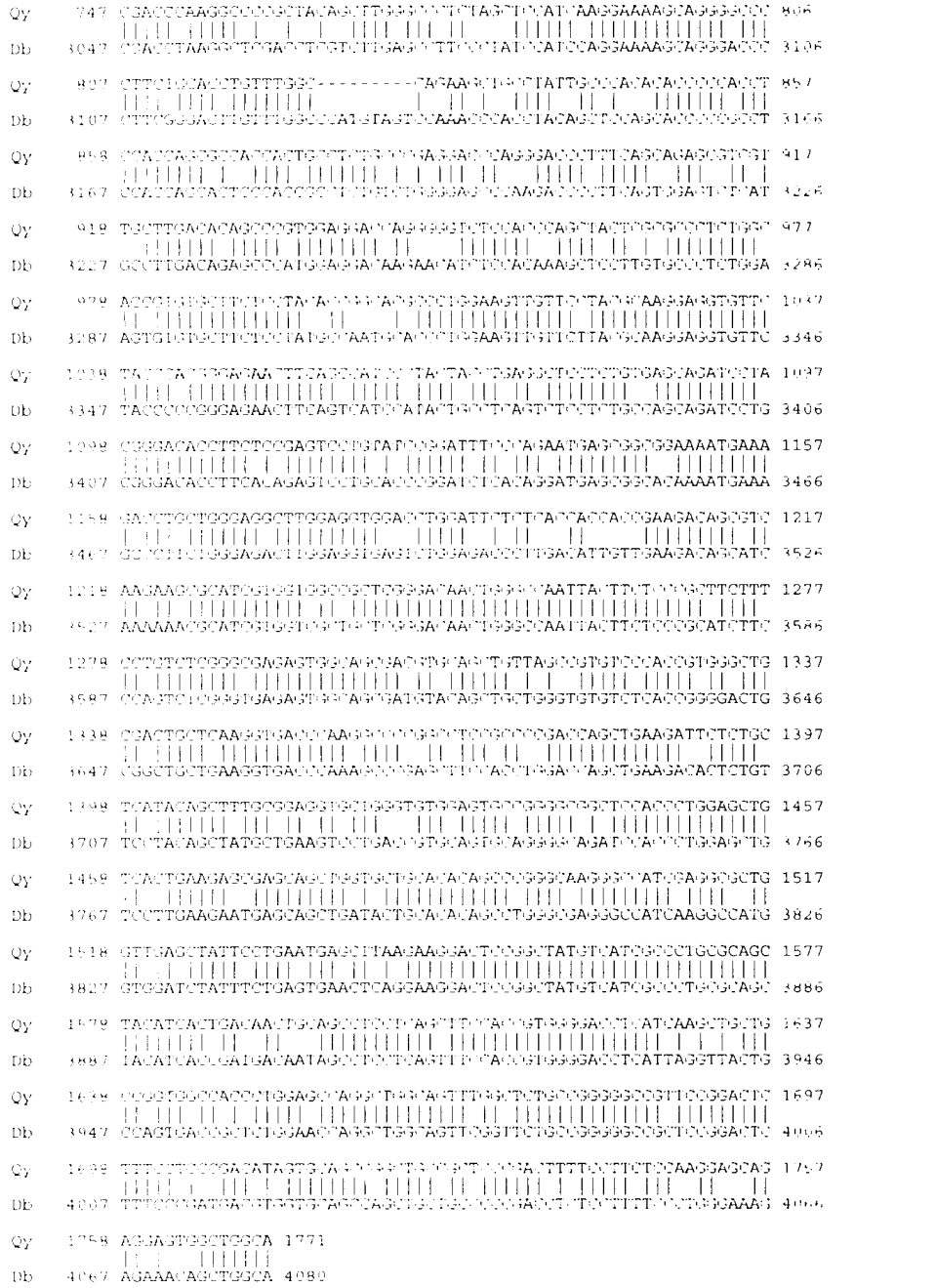
Query Match 25.2%; Score 954.4; DB 9; Length 1247;
Best Local Similarity 99.7%; Pred. No. 6.4e-187;
Matches 955; Conservative 1; Mismatches 2; Indels 0; Gaps 0;



RESULT 6
US-09-803-126-3
; Sequence 3, Application US/09803126
; Patent No. US20020093190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Pong, Gary T.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4375
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse myosin related protein (MRP) variant 2 cDNA
US-09-803-126-3

Query Match 25.0%; Score 946.4, DB 10, Length 4375;
Best Local Similarity 72.9%; Pred. No. 4.1e-185;
Matches 1322; Conservative 0, Mismatches 416, Indels 76; Gaps 5;





```
RESULT 7
US-09-764-868-1347/c
; Sequence 1347, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ32
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1347
;   LENGTH: 24757
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-868-1347

  Query Match             11.7%;  Score 442;  DB 9;  Length 24757;
  Best Local Similarity   92.1%;  Pred. No. 2.1e-81;
  Matches  466;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy    3254 GGCCTGGAAGTCAACTATGGCTCAGCTGACAACCCCAGAGCAACTGTTTTGAGCTGTA 3313
           |||| ||       | | ||     ||||   | |||       | | ||  |
Db    4289 GGCCGGGGCAGAGGCAGGGGCTGGCTGGGGAGCCCTCTTGTAACAGCCTTTCCTGTT 4348

Qy    3314 CAGGCCAGGAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACT 3373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4229 CAGGCCAGGAGCTGCTATACACCACTGTCTTCCTGATAGACAGCAGTGCCTCTTGCACT 4170

Qy    3374 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCTCCT 3433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4169 GAGTGGCCCAGCATCAACTGAGAGGAGTGCAGGCCGGGGAGAGAAGAGGATGAGGCTCCT 4110

Qy    3434 CCGGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 3493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4109 CCGGGCCCAAGTCTCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAA 4050

Qy    3494 CCTGCCTCAGCACAGCCCAGCGGGGCCACATCTAGGCCATGAGGCAGGGCTGCTATCAC 3553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4049 CCTGCCTCAGCACAGCCCAGCGGGGCCACATCTAGGCCATGAGGCAGGGCTGCTATCAC 3990

Qy    3554 GTCACCAGCAGGCAAAGAAAACAGCCAGACCTCTCCAGGAGGGCCTGGGGCCAAAGCGG 3613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3989 GTCACCAGCAGGCAAAGAAAACAGCCAGACCTCTCCAGGAGGGCCTGGGGCCAAAGCGG 3930

Qy    3614 GCTGCAGGAACTGGGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGAGATGCCCACCCGA 3673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3929 GCTGCAGGAACTGGGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGAGATGCCCACCCGA 3870

Qy    3674 CCCCAGGCTCCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGAGAAACAGCTGCT 3733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3869 CCCCAGGCTCCGCCCAGGCCCCACATTAGCACAAGCCCAGGCATGGAGAAACAGCTGCT 3810

Qy    3734 GAGGAAATAAAACTCCCTAAAAAAAA 3759
           ||||||||||||||||| |||| |
Db    3809 GAGGAAATAAAACTCCCTGGAGAAAA 3784


RESULT 8
US-09-954-456-176/c
; Sequence 176, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti Cancer Therapeutic Agents Using
;  TITLE OF INVENTION: Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE: 2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
```

```
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 176
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-176

  Query Match             7.7%;  Score 292;  DB 10;  Length 311;
  Best Local Similarity   98.1%;  Pred. No. 3.4e-51;
  Matches  306;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy    3447 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 3506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 252

Qy    3507 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCACGTCACCAGCAGGC 3566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCACGTCACCAGCAGGC 192

Qy    3567 AAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 3626
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     191 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 132

Qy    3627 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCGACCCCACGTCCGC 3686
           |||||||||||||||||||||||||||||||||||||||||||||||||| | |||||
Db     131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCGACCCCACG-TCCGC 73

Qy    3687 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 3746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 13

Qy    3747 TCCCTAAAAAAA 3758
           ||||| | |||
Db      12 TCCCTGGAGAAA 1


RESULT 9
US-09-954-456-861/c
; Sequence 861, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 861
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-456-861

  Query Match             7.7%;  Score 292;  DB 10;  Length 311;
  Best Local Similarity   98.1%;  Pred. No. 3.4e-51;
  Matches  306;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy    3447 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 3506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 TCACCCACATGGTCTGCCTTGGATGCTATCAGATCACTGTTCTAGAACCTGCCTCAGCAC 252

Qy    3507 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCACGTCACCAGCAGGC 3566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 AGCCCAGCCGGCCCACATGCAGGCCATGAGGCAGGGGCTGCTATCACGTCACCAGCAGGC 192

Qy    3567 AAAGAAAACAGCCAGACCCTCTCCAGGACGGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 3626
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     191 AAAGAAAACAGCCAGACCCTCTCCAGGACTGCCTGGGGCCAAAGCGGGCTGCAGGAACTC 132

Qy    3627 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCGACCCCACGTCCGC 3686
           |||||||||||||||||||||||||||||||||||||||||||||||||| | |||||
Db     131 GGCTGGGGCACCTGAGGTTGCCCAGTCTGAGGGAGATGCCCACCGACCCCACG-TCCGC 73

Qy    3687 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 3746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 CCAGGCCCCACATTAGCACAAGCCCAGGCATGGGAGAAACAGCTGCTGAGGAAATAAAAC 13

Qy    3747 TCCCTAAAAAAA 3758
           ||||| | |||
Db      12 TCCCTGGAGAAA 1


RESULT 10
US-09-880-107-1520/c
; Sequence 1520, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1520
```

```
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA121277
US-09-880-107-1529

  Query Match              7.7%;  Score 292,  DB 19,  Length 311,
  Best Local Similarity   98.1%;  Pred. No. 3.4e-51,
  Matches  ...,  Conservative    0,  Mismatches    5,  Indels    1,  Gaps    1,

Qy    3447 T...  3506
           ...  ...
Db     311 TCA...  262

Qy    3507 A...  3566
           ...  ...
Db     261 A...  192

Qy    3567 AAA...  3626
           ...  ...
Db     191 AAA...  132

Qy    3627 GG...  3686
           ...  ...
Db     131 GG...  74

Qy    3687 C...  3746
           ...  ...
Db      73 CA...  13

Qy    3747 TC...  3758
           ...  ...
Db      12 TC...  1

RESULT 11
US-09-864-761-7087/c
; Sequence 7087, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharron G.
;   APPLICANT:  Rank, David R.
;   APPLICANT:  Hanzel, David K.
;   APPLICANT:  Chen, Wensheng
;   TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;   TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE: Aeomica-X-1
;   CURRENT APPLICATION NUMBER: US/09/864,761
;   CURRENT FILING DATE:  2001-05-23
;   PRIOR APPLICATION NUMBER: US 60/180,312
;   PRIOR FILING DATE: 2000-02-04
;   PRIOR APPLICATION NUMBER: US 60/207,456
;   PRIOR FILING DATE: 2000-09-26
;   PRIOR APPLICATION NUMBER: US 09/632,366
;   PRIOR FILING DATE: 2000-08-03
;   PRIOR APPLICATION NUMBER: GB 24263.6
;   PRIOR FILING DATE: 2000-10-04
;   PRIOR APPLICATION NUMBER: US 60/236,359
;   PRIOR FILING DATE: 2000-09-27
;   PRIOR APPLICATION NUMBER: PCT/US01/00666
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00662
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00661
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00670
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: US 09/234,687
;   PRIOR FILING DATE: 2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,204
;   PRIOR FILING DATE: 2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7087
;   LENGTH: 526
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
US-09-864-761-7087

  Query Match              4.6%;  Score 174;  DB 10;  Length 526;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-27;
  Matches  174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1656 CCAGGCTGGCAGTTTGGCTCTGCCGGGAGCCGTTCCGGACTCTTTCCTGCCGACATATTG 1715
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 CCAGGCTGGCAGTTTGGCTCTGCCGGGAGCCGTTCCGGACTCTTTCCTGCCGACATATTG 137

Qy    1716 CAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGAAGCAGACGAGAGCCCCTGGCAAG 1775
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 CAGCCGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGAAGCAGACGAGAGCCCCTGGCAAG 77

Qy    1776 GGTCAGCTGTCCAACGGGGAAGCAGGGCTGGCTCCGTGGGAGAGGTCCTCAGAG 1829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 GGTCAGCTGTCCAACGGGGAAGCAGGGCTGGCTCCGTGGGAGAGGTCCTCAGAG 23

RESULT 12
US-09-864-761-23818/c
; Sequence 23818, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
```

; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23818
; LENGTH: 170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EST_HUMAN HIT: AV649878.1, EVALUE 5.00e-91
; OTHER INFORMATION: SWISSPROT HIT: P05895, EVALUE 1.00e+00
; OTHER INFORMATION: NT HIT: gi11426247, EVALUE 1.50e+00
US-09-864-761-23818

Query Match 4.5%; Score 170; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 3e-26;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1650 GCTGGCAGTTTGGTCTCTGGCGGGGGGGTCGTTCGGGACTCTTTCCTGCGGACATAGTGCAGC 1719
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 170 GCTGGCAGTTTGGTCTCTGGCGGGGGGGTCGTTCGGGACTCTTTCCTGCGGACATAGTGCAGC 111

Qy 1720 CGGCTGCCGCTCCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTC 1779
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110 CGGCTGCCGCTCCCGACTTTTCCTTCTCCAAGGAGCAGAGGAGTGGCTGGCACAAGGGTC 51

Qy 1780 AGCTGTCCAACGGGGAACCCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAG 1829
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 50 AGCTGTCCAACGGGGAACCCAGGGCTGGCTCGGTGGGACAGGGCCTCAGAG 1

RESULT 13
US-09-864-761-24717/c
; Sequence 24717, Application US/09864761
; Patent No. US20020049363A1
; GENERAL INFORMATION:
; APPLICANT: Fenn, Sharron G.
; APPLICANT: Rank, David R
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/190,312
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24717
; LENGTH: 233
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: SWISSPROT HIT: P70085, EVALUE 2.90e-01
; OTHER INFORMATION: NT HIT: AF004636.1, EVALUE 8.70e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW753302.1, EVALUE 9.00e-78
US-09-864-761-24717

Query Match 3.6%; Score 137.8; DB 10; Length 233;
Best Local Similarity 78.7%; Pred. No. 1.3e-19;
Matches 177; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

Qy 2647 GCCAGCTGGTGCCGCCCCTGCAGCCCGCCGAATACCTCAACAGCGTCGTAGTGGACCTAGG 2706
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 232 GCCAGCTGGTGCCGCCCCTGCAGCCCGCCGAATACCTCAACAGCGTCGTAGTGGACCTAGG 174

Qy 2707 ACGTGAGCCTGCACAG-CGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCA 2765
|||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db 173 ACGTGAGCCTGCACAGCCGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCA 114

Qy 2766 CCTACATCAGCACCCACTACAGCCACGTCGTCGTGGACTACCTTCAGGGGAAGCTGCCAG 2825
|||||||||||||||||||||||||||||| | | || | |||||
Db 113 CCTACATCAGCACCCACTACAGCCACGTCATCGTCGACTACTTCGGCGAGCTGCCAG 54

```
Qy    2826 TCAGGGCAAGGCAGAGGTGTAGTTGGCAGGCTGGGGGGCCTGT 2870
           |||| ||| |  | | | ||| | ||| | |||| |||| |||
Db      53 TTGAGGCGGGGCTTCCTGCTGGGGGACCTCCAAGCCCCTCTCC 9


RESULT 14
US-09-864-761-8013/c
; Sequence 8013, Application US/09864761
; Patent No. US20020048764A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wencheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00671
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/818,409
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8013
;   LENGTH: 566
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC011014.2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  3.5
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  2.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  2.6
US-09-864-761-8013
```

```
  Query Match              3.6%;  Score 147.8;  DB 10;  Length 566;
  Best Local Similarity    86.7%;  Pred. No. 1.7e-15;
  Matches  177;  Conservative    0;  Mismatches   47;  Indels    1;  Gaps    1;

Qy    2647 GCCAGCTGGTGCGGGCCCCTGCAGCCCGCCGAATACCTCAACAAGCGTGATAGTGGACTAGT 2706
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 GCCAGCTGGTGCGGGCCCCTGCAGCCCGCCGAATACCTCAACAAGCGTGATAGTGGACTAGT 438

Qy    2707 ACGTGAGCCTGCACAGCGGGGGGCTCCACTGGGAGACCCAACCCACTTTGATAACTCCA 2766
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACGTGAGCCTGCACAGCGGGGGGCTCCACTGGGAGACCCAACCCACTTTGATAACTCCA 260

Qy    2767 CCTACATTACGCACCCACTACAGCCACGCCTGCCTGCCTCCGAGCTGTCCACTGCTGCACT 2826
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 CCTACATCACGCACCCACTACAGCCACGCCTGCCTGCCTCCGAGCTGTCCACTGCTGCACT 200

Qy    2827 TCAGGGCAAGGCAGAGGTGTAGTTGGCAGGCTGGGGGGCCTGT 2870
           |||| ||| |  | | | ||| | ||| | |||| |||| |||
Db     199 TTGAGGCGGGGCTTCCTGCTGGGGGACCTCCAAGCCCCTCTCC 156


RESULT 15
US-09-783-590-1780
; Sequence 1780, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products Th...
;  FILE REFERENCE: PG-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2001-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1780
;   LENGTH: 312
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (20)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (26)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (42)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (123)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (141)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (142)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (159)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (165)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (168)
;   OTHER INFORMATION: n equals a,t,g, or c
```

; NAME/KEY: misc feature
; LOCATION: (169)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (170)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (181)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (196)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (228)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (244)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (290)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1780

Query Match 3.0%; Score 113.8; DB 10; Length 312;
Best Local Similarity 80.4%; Pred No 1.2e-14;
Matches 234; Conservative 0; Mismatches 43; Indels 14; Gaps 9;

```
Qy    2699 GGACCAGGACGTGAGCCTGCACAG-CGGCGGCTCCACTGGG-AGACCCCACTGCACTTCG 2756
           |||||||||||| ||||| ||||| ||||||||| |||||| ||||||||||||||||||
Db       8 GGACCAGGAGGTNAGCCTNCACAGCCGGGGGCTCNACTGGGAAGACCCCACTGCACTTCG 67

Qy    2757 ATAACTCCACCTACATCAGCACCCACTACAGCCAGGTGCTGTGGGACTACCTT---CAGG 2813
           ||||||||||||||||||||||||||||||||||||||||||||||||||    | |     ||
Db      68 ATAACTCCACCTACATCAGTACCCACTACAGCCAGGTGCTGTGGGAACTACCTTGNAGGG 127

Qy    2814 GGAAGCTGCCAGT-CAGCGCCAAGGCAGACGCGCAGCTCGCCAGGCT--GGCCGCCCTGC 2870
           | |||||||||||  ||||||||||||||| |     ||    | |  | |  ||  ||
Db     128 GTAAGCTGCCAGTNNAGCGCCAAGCCAGACGNGACATNCGNNNAGGTTGGCCGNCCTGGC 187

Qy    2871 AGCACCT CAGCAAGGCTAACAGG--AATAC C CTCAGGGCAGG--ACCTGCTAGCTTA 2925
           |||||||  |||||||||||||||  | |||||||||||| ||||    |     || |||
Db     188 AGCACCTGNAGCAAGGCCAACAGGGAATTACCCCCTCAGGNCAGGGACCTGCTAAGNTTA 247

Qy    2926 CGTGCCAAAGCAGCTGCAACGGCAGGTG-AACACGGCCTCCATCAAGAACC 2975
           |||||||||||||||| ||||| ||||| ||||||||||||| |||| | |
Db     248 CGTGCCAAAGCAGCTGNAACGGNAGGTGNAACACGGCCTCCANCAAGGAAC 298
```

Search completed: January 10, 2003, 02:23:40
Job time : 275.541 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:10:54 ; Search time 179.767 Seconds
(without alignments)
10795.718 Million cell updates/sec

Title: US-09-803-126-2
Perfect score: 6293
Sequence: 1 [illegible] 6293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	55	0.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	46.8	0.7	364	4	US-08-905-223-162	Sequence 162, App
	3	44.8	0.7	776	4	US-09-535-008-36	Sequence 36, Appl
c	4	43	0.7	1357	4	US-09-669-680-3	Sequence 3, Appli
	5	43	0.7	2744	3	US-09-071-101-1	Sequence 1, Appli
	6	43	0.7	2744	3	US-09-369-618-1	Sequence 1, Appli
	7	43	0.7	2744	3	US-09-369-617-1	Sequence 1, Appli
c	8	42.4	0.7	71300	4	US-09-443-501A-2	Sequence 2, Appli
	9	41.6	0.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	10	41	0.7	3483	2	US-08-728-323A-1	Sequence 1, Appli
	11	41	0.7	3489	4	US-09-298-568-1	Sequence 1, Appli
c	12	41	0.7	32007	3	US-08-770-378-20	Sequence 20, Appl
c	13	41	0.7	32007	4	US-08-757-669A-20	Sequence 20, Appl
c	14	41	0.7	32007	4	US-09-230-371A-20	Sequence 20, Appl
c	15	40.8	0.6	68750	3	US-09-335-409-1	Sequence 1, Appli
c	16	40.8	0.6	68750	4	US-09-568-102-1	Sequence 1, Appli
c	17	40.8	0.6	68750	4	US-09-567-969-1	Sequence 1, Appli
c	18	40.8	0.6	68750	4	US-09-568-480-1	Sequence 1, Appli
c	19	40.8	0.6	68750	4	US-09-568-486-1	Sequence 1, Appli
c	20	40.8	0.6	68750	4	US-09-568-472-1	Sequence 1, Appli
c	21	40.8	0.6	68750	4	US-09-567-899-1	Sequence 1, Appli
	22	40	0.6	2074	4	US-08-630-915A-19	Sequence 19, Appl
	23	40	0.6	2269	4	US-09-394-645-1	Sequence 1, Appli
	24	40	0.6	2269	4	US-09-243-560B-1	Sequence 1, Appli
c	25	40	0.6	3251	4	US-09-085-199B-6	Sequence 6, Appli
	26	39.8	0.6	1150	1	US-08-161-406-1	Sequence 1, Appli
	27	39.8	0.6	2943	4	US-09-404-879A-385	Sequence 385, App
	28	39.8	0.6	3124	4	US-09-734-030-1	Sequence 1, Appli
	29	39.6	0.6	2984	5	PCT-US93-00893-1	Sequence 1, Appli
c	30	39.6	0.6	2984	5	PCT-US93-00893-2	Sequence 2, Appli
	31	39.4	0.6	7286	3	US-09-331-581-3	Sequence 3, Appli
	32	39.4	0.6	7938	3	US-09-331-581-14	Sequence 14, Appl
c	33	39.4	0.6	51259	3	US-08-781-891-209	Sequence 209, App
	34	39.2	0.6	1068	4	US-09-077-675A-6	Sequence 6, Appli
	35	39.2	0.6	1122	4	US-09-077-675A-9	Sequence 9, Appli
	36	38.8	0.6	280	4	US-08-667-005-17	Sequence 17, Appl
	37	38.8	0.6	289	4	US-09-244-796-17	Sequence 17, Appl
	38	38.8	0.6	477	4	US-09-135-994-1	Sequence 1, Appli
c	39	38.8	0.6	4741	1	US-07-695-472B-4	Sequence 4, Appli
c	40	38.8	0.6	4742	1	US-08-250-740-35	Sequence 35, Appl
c	41	38.6	0.6	954	3	US-08-911-863-24	Sequence 24, Appl
c	42	38.6	0.6	954	4	US-08-479-453-24	Sequence 24, Appl
c	43	38.6	0.6	954	4	US-09-479-453-24	Sequence 24, Appl
	44	38.6	0.6	1576	1	US-08-157-101A-6	Sequence 6, Appli
c	45	38.6	0.6	16442	3	US-08-781-891-208	Sequence 208, App

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT POWDERY MILDEW VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,813
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION
; NAME: BENT, Stephen A
; REGISTRATION NUMBER 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pT2gpt-F1s
US-08-232-463-14

Query Match 0.9%; Score 55; DB 1; Length 7218;

```
  Best Local Similarity   4.0%;  Pred. No. 0.0014;
  Matches   16;  Conservative   22;  Mismatches  158;  Indels    0;  Gaps    0;

Qy    2381 AAAAAAAAGCAGAAATAACTAGAGTCTTTGGGTGTTGGCTTGAGAGAGGACACCCCA 2440
           |||||||||  ||||| ||   ||||     ||| |  ||   ||||||||| |    |
Db    1438 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1377

Qy    2441 AAAAAAATTGAAAGCAACCCCTCAGCGGCCCAACAGCTTCCAACAGAAAGGGGACTATTTC 2500
           |||||  ||||| |||    || |    |||||     || ||||| ||||  |    |
Db    1376 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316

Qy    2501 CCTAAGATCTGCCAAGATCTATAGAGTGAAGAGCTTCAAACCTTCCATCCAAGTTTCAG 2560
           ||||| |||| ||||   || |  |||| ||||| || ||||    ||  |||| |  ||
Db    1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256

Qy    2561 ATTCCCAAGAGCAGAAGCAAGAGCGGAGGAGGAGGAGGATGAGACGGCCGAGTTGTCC 2620
           |  |||||||| |||||||| |||||||||||||||| |||| |  |||  |  |||
Db    1255 RRRRRRRYRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196

Qy    2621 CTCCTCTCCCCCCTCCCCGGTTGTGAAGAAGCCGCTGAAGGCAAGCAGGCCCAAAGCC 2680
           ||  | |||||  |  ||| |  ||||| ||| |||   |||||||||||||||||||
Db    1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136

Qy    2681 CTAAAGGAAGATCAGCCAGAGCCCGCCTAGGAGGAAGTACGACCCAGGGCGAGAACCCC 2740
           | |||||||| |||| ||||   |  |||||||| |  ||   |||| |||||
Db    1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076

Qy    2741 CCGGTGCACAGCTCCAACTCCGCACCTCAGCACCCCA 2777
           || | | |  | | | |||.    | ||| ||| |
Db    1075 RRRRRRRRRATGGCAAGCTCCCTCGAGCTGCAGCCAA 1039


RESULT 2
US-08-905-223-162
; Sequence 162, Application US/08905223
; Patent No. 6221329
; GENERAL INFORMATION:
;   APPLICANT: Edwards, Jean-Baptiste D.
;   APPLICANT: Corbett, Aymeric
;   APPLICANT: Lacroix, Bruno
;   TITLE OF INVENTION: ST ECTO FOR SECRETED PROTEINS
;   NUMBER OF SEQUENCES: 503
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Knobbe, Martens, Olson & Bear
;     STREET: 501 West Broadway
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92101 3505
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy Disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: Win95
;     SOFTWARE: Word
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/905,223
;     FILING DATE:
;     CLASSIFICATION: 536
;   ATTORNEY AGENT INFORMATION:
;     NAME: Israelsen, Ned A.
;     REGISTRATION NUMBER: 29,655
;     REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION
;     TELEPHONE: (619) 235-8550
;     TELEFAX: (619) 235-8176
; INFORMATION FOR SEQ ID NO: 162:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 564 base pairs
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: DOUBLE
;     TOPOLOGY: LINEAR
;   MOLECULE TYPE: cDNA
;   ORIGINAL SOURCE:
;     ORGANISM: Homo Sapiens
;     TISSUE TYPE: Brain
;   FEATURE:
;     NAME/KEY: sig_peptide
;     LOCATION: 56..1271
;     IDENTIFICATION METHOD: Von Heijne matrix
;     OTHER INFORMATION: score 3.6
;     OTHER INFORMATION: seq CTSLLQLYDASNS/EW
US-08-905-223 162

  Query Match              0.7%;  Score 46.8;  DB 4;  Length 564;
  Best Local Similarity   54.5%;  Pred. No. 0.052;
  Matches   87;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    ;

Qy     271 TCCCACCCCTGGAAAGATCCCAGGCCTCTTGGACGTGGGGCATGTGCTAGACATGCTGCC 330
           ||  |  | || ;|| ||| ||  || ||| || |  |  |  |  | ||  |
Db      40 TTCAGCTCCAGGCTAGATGCCAGACCAGTTTGACCAGGGGGTCTGCTGAACCAGCTGGC 99

Qy     331 TCAGGCTGGCATCCTGGAGATCATAGGCACCCGGAGTACCCACTTCCCCTGCGAGTCTC 390
           |  | || || |||||||  |  | | || |  |  |  |  |  |  |  |  |
Db     100 GTACTCAGGGATGCTGGAGACTGTGAGAATCCGCAAAGCTGGCTATGCGTCCGAAGACC 159

Qy     391 CTTCCAAGTCTTTCTGGCAAGGTTCCATGCCCTG 424
           ||| | | ||||     |||||  | |  |||
Db     160 CTTCCAGGACTTTCATAAAAGGTATAAAGCTGCTG 193


RESULT 3
US-09-535-008-36
; Sequence 36, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
;  APPLICANT: Wong, Alexander K.C.
;  APPLICANT: Tavtigian, Sean V.
;  APPLICANT: Teng, David H.-F.
;  TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
;  TITLE OF INVENTION: AND OTHER CANCER TYPES
;  FILE REFERENCE: 2318-259
;  CURRENT APPLICATION NUMBER: US/09/535,008
;  CURRENT FILING DATE: 2000-03-23
;  EARLIER APPLICATION NUMBER: U.S. 60/126,806
;  EARLIER FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;  LENGTH: 776
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-535-008-36

  Query Match              0.7%;  Score 44.8;  DB 4;  Length 776;
  Best Local Similarity   47.0%;  Pred. No. 0.15;
  Matches  139;  Conservative    0;  Mismatches  157;  Indels    9;  Gaps    ;

Qy    1075 GCGGTTAGATGGCGAGAACCCTCAGCAGGTCTGGAGATAACAAGGGAATCTGCGGTT 1134
           || |  |  |||  |  | |           |    |  |  |  |  |  |  |
Db     268 GCAGTGCGGCGGGCTTGTCCTCTTCCTCCTACAGCTGCGCAGCAGGGGGGGTGTGCAT 327

Qy    1135 CCTGGGGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGCATCGGGCATCTCCTGGTCA 1194
           | | ||||  | | ||| |: |    |  |  :  :  :  :  :  :  :  :
Db     328 GCGGAGGGACACAGGGCTGGAGACAGCCCTCAATGCTAAGG-CTATAAGGGTAGCAACT 386

Qy    1195 GCAGGCCCAGCGACGGCGGCACTCCGAGATGACTCTTTAGTCATTTTTGCCTAGTT 1254
           ||| ||| |||  | |  |  ||  ||  ||||           |  | ||
Db     387 CCAGTCCCTGGGGCAGGGCCCCATACTCAGAAGCTGGAGAATCATCAGAAGATGCAGCA 446

Qy    1255 GTGGCGCAACCCAGATGAGCAACAGAAICAGCCGGGTGGGCCCTAATTTTGATCTGTT 1314
           |  |||||| |      || | ||  |    ||  |  | | |  |  | | | ||
Db     447 GGAGCGCAAGCGCCGGAGAAGCACCAGTATCCCCCCGGG-  AAGCCCTTGTATTT 97

Qy    1315 CAGCTCCTTGCTCCCACACTGCCCTGCAGAAACACATCTAAAT...TATTT 1370
           |  | | | ||| ||   | |: ?| |  |  ||||| |  :      :  | ||
Db     506 CGCCCACCTTGGCTGCCCTCTGCTACCCCTGCCACCCC.TTCCTTTTTCTTT 556
```

RESULT 4
US-09-668-680-3/c
; Sequence 3, Application US/09668680
; Patent No. 6436703
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Xu, Chongjun
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: N: 6436703el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 790CIP2A
; CURRENT APPLICATION NUMBER: US/09/668,680
; CURRENT FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 3
; LENGTH: 1357
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (284)..(904)
US-09-668-680-3

Query Match 0.7%; Score 43; DB 4; Length 1357;
Best Local Similarity 52.5%; Pred. No. 0.54;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

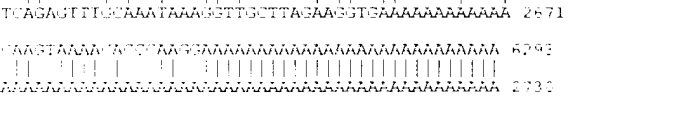
RESULT 5
US-09-071-101-1
; Sequence 1, Application US/09071101
; Patent No. 6013503
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,101
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E.
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 61..2325
; OTHER INFORMATION
US-09-071-101-1

Query Match 0.7%; Score 43; DB 3; Length 2744;
Best Local Similarity 52.5%; Pred. No. 0.78;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

RESULT 6
US-09-369-618-1
; Sequence 1, Application US/09369618
; Patent No. 6100041
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Jaspers, Stephen R
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; FILE REFERENCE: 97-05D2
; CURRENT APPLICATION NUMBER: US/09/369,618
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: US 09/071,101
; EARLIER FILING DATE: 1998-05-01
; EARLIER APPLICATION NUMBER: US 60/044,015
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2744
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)...(2325)
US-09-369-618-1

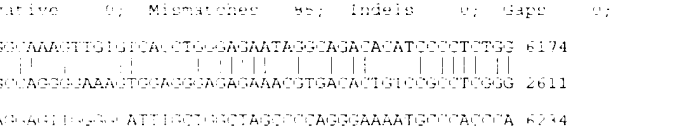
Query Match 0.7%; Score 43; DB 3; Length 2744;

Best Local Similarity 52.5%; Pred. No. 0.78;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;



RESULT 7
US-09-369-617-1
; Sequence 1, Application US/09369617
; Patent No. 6127162
; GENERAL INFORMATION:
; APPLICANT: LOH, Si
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; FILE REFERENCE: 97-05D1
; CURRENT APPLICATION NUMBER: US/09/369,617
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: US 09/071,101
; EARLIER FILING DATE: 1998-05-01
; EARLIER APPLICATION NUMBER: US 60/044,015
; EARLIER FILING DATE: 1998-05-30
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2744
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 261...(2325)
US-09-369-617-1

Query Match 0.7%; Score 43; DB 3; Length 2744;
Best Local Similarity 52.5%; Pred. No. 0.78;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;



RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6703342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Fulton, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Kinmann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 40062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/140,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71389
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 0.7%; Score 42.4; DB 4; Length 71389;
Best Local Similarity 47.1%; Pred. No. 5.8;
Matches 164; Conservative 0; Mismatches 181; Indels 4; Gaps 1;

RESULT 9
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899119
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTT3EF Pis
US-08-232-463-14

Query Match 0.7%; Score 41.6; DB 1; Length 7218;
Best Local Similarity 2.1%; Pred. No. 2.8;
Matches 8; Conservative 21; Mismatches 159; Indels 0; Gaps 0;

Qy 3086 TTTCAGAAAAGTAATTAGTTCCTTGAGCTAAGGTTCGACCTCGTCTTGAGCTTCCCTATCC 3085
Db 1067 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1126

Qy 3084 ATTTAGGAAAGTAATTAGTTCCTTGGGACTTCGTTTGGGCCATGTAGTCCAAAGCCAGCT 3145
Db 1127 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1186

Qy 3146 ACAGTTCCATCAGTTTTATTCTATCATGATTCCTTCTTTTGGGTAGCCCAAG 3205
Db 1187 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1246

Qy 3206 ATTTTTTAATTGAATTTATGCTTTGATATAATCATGAGGACAAGAACATCTCCACA 3265
Db 1247 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1306

Qy 3266 AAGTTTTTTGTGTTTTTGAAGTTTTGTTTTCTATGCAATGCATCTTGAATTG 3325
Db 1307 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1366

Qy 3326 TTTTAGTTAATGAATTTTTATGTTTAATGAATTTCAGTCATCATATTGTTCAGT 3385
Db 1367 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1426

Qy 3386 TTCTTTTTTAGTAATTTTTG 3405
Db 1427 YYYYYYYYYYAATAAATTTTG 1446

RESULT 10
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Rabensky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

Query Match 0.7%; Score 41; DB 2; Length 3489;
Best Local Similarity 47.2%; Pred. No. 2.8;
Matches 126; Conservative 6; Mismatches 141; Indels 0; Gaps 0;

Qy 2484 AGAAACGGGACTATTTCCAGAAGATGGAAAAATCCCCATAAATTAAAAACCTTTAAAC 2543
Db 2021 AGCAGCAGTAGAATGAGCAGCAGTAGGATGAGTAGTAATAAATGAATGAGTAGTAGTATG 2080

Qy 2544 CTCCAGTAAAGTTCAGATCCCCAAAGGAGATGGAGAAAAAAAAAAAAAAAAAGAGC 2603
Db 2081 AGCAGCAGCAGGATGAGCAGCAGTAGTAGTAGATGAGTAGTAGTAGCAGCAGTATG 2140

Qy 2604 AGACCCCCGAGTTCTCCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 2663
Db 2141 AGCAGGATGAGCAGCAGCAGGATGAGTAGTAGCAGCAGTAATGAGTAGTAGCAGCAGGATG 2200

Qy 2664 CAAGCAGCCCCTAAAGCCCTAAAGAAAGAATGAGTAGTAGTAGTAGTAGTAGCAGTAGCTA 2723
Db 2201 AGCAGCAGCAGCAGGATGAGCAGCAGTAGTAGTAGGATGAGTAGTAATAAATGAATGAGCAGG 2260

Qy 2724 CCCAGGGAGAGGATCCCCCGGTCCA 2748
Db 2261 AGCAGCAGGAGGAGCAGGAGCAGCA 2285

RESULT 11
US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHABDO VIRUS DNA ACTS IN TRANS ON A UNIT OF PRADING
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422

```
;  EARLIER FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;    LENGTH: 3489
;    TYPE: DNA
;    ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-Sub-1

  Query Match                 0.7%;  Score 41;  DB 4;  Length 3489;
  Best Local Similarity   47.2%;  Pred. No. 2.8;
  Matches  126;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy    2484 AGAAACGGGACTATTTCAGAAGATGGGGCAAGATCGGATCAGAGTGAAGACGGTGAAAC 2543
Db    2021 AGTAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 2080

Qy    2544 CTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGAGGATG 2603
Db    2081 AGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGC 2140

Qy    2604 AGACGGCGAGTTCTCCCCTCCTCCCCCCTCCCCCTTTCTGAAGAAGCCGCTGAAGG 2663
Db    2141 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATG 2200

Qy    2664 CAAGCAGGCCAAAGCCGTAAAGGAAGATGAGGTAGACGCCGCCAGGAGGAAGTACCGA 2723
Db    2201 AGTAGCAGCAGGATGAGCAGCAGTAGCAGCAGCAGGATGAGTAGTAGCAGTAGGATGAATAGG 2260

Qy    2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
Db    2261 AGTAGCAGCAGGAGCAGGAGCAGCA 2285


RESULT 12
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5840564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  32342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20

  Query Match                 0.7%;  Score 41;  DB 1;  Length 32207;
  Best Local Similarity   47.2%;  Pred. No. 8.6;
  Matches  126;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy    2484 AGAAACGGGACTATTTCAGAAGATGGGGCAAGATCGGATCAGAGTGAAGACGGTGAAAC 2543
Db   19976 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 20035

Qy    2544 CTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGAGGATG 2603
Db   19916 AGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGC 19975

Qy    2604 AGACGGCGAGTTCTCCCCTCCTCCCCCCTCCCCCTTTCTGAAGAAGCCGCTGAAGG 2663
Db   19856 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATG 19915

Qy    2664 CAAGCAGGCCAAAGCCGTAAAGGAAGATGAGGTAGACGCCGCCAGGAGGAAGTACCGA 2723
Db   19796 AGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAATAGG 19855

Qy    2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
Db   19736 AGCAGCAGGAGGAGCAGGAGCAGCA 19712


RESULT 13
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED HERPESVIRUS SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,669A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  49185-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-20

  Query Match              0.7%;  Score 41;  DB 4;  Length 32207;
  Best Local Similarity    47.2%;  Pred. No. 8.6;
  Matches  125;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy    2494 AGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACGGTGAAAC 2543
Db   19976 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 19917

Qy    2544 CTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGATG 2603
Db   19916 AGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGC 19857

Qy    2604 AGACCCCCGACTTGTCCCCTCTTCTTCCCCTTCCCCGGTTGTGAAGAAGCCGCTGAAGG 2663
Db   19856 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATG 19797

Qy    2664 CAAGCAGGCCCAAAGCCGTAAAGGAAGATGAGGCAGAGCCGCCCCAGGAGGAAGTACCGA 2723
Db   19796 AGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAACAGG 19737

Qy    2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
Db   19736 AGCAGCAGGAGGAGCAGGAGCAGCA 19712

RESULT 14
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. [illegible]
; GENERAL INFORMATION:
;  APPLICANT: [illegible]
;  APPLICANT: Bohenzky, Roy A
;  APPLICANT: [illegible], James J
;  APPLICANT: Edelman, Isidore S
;  APPLICANT: [illegible], Patrick S
;  TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: 45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/[illegible],371A
;  CURRENT FILING DATE: 1999-11-17
;  PRIOR APPLICATION NUMBER: PCT/US97/13346
;  PRIOR FILING DATE: 1997-07-22
;  NUMBER OF SEQ ID NOS: [illegible]
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 32207
;   TYPE: DNA
;   ORGANISM: Kaposi's sarcoma associated herpesvirus
US-09-230-371A-20

  Query Match              0.7%;  Score 41;  DB 4;  Length 32207;
  Best Local Similarity    47.2%;  Pred. No. 8.6;
  Matches  125;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy    2494 AGAAACGGGACTATTTCCAGAAGATGGGGCAAGATCCGATCAGAGTGAAGACGGTGAAAC 2543
Db   19976 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATG 19917

Qy    2544 CTCCAGCCAAGGTTCAGATCCCCCAAGAGGAGATGGAGGAGACGGAGGAGGAGGAGGATG 2603
Db   19916 AGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGC 19857

Qy    2604 AGACCCCCGACTTGTCCCCTCTTCTTCCCCTTCCCCGGTTGTGAAGAAGCCGCTGAAGG 2663
Db   19856 AGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATG 19797

Qy    2664 CAAGCAGGCCCAAAGCCGTAAAGGAAGATGAGGCAGAGCCGCCCCAGGAGGAAGTACCGA 2723
Db   19796 AGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAACAGG 19737

Qy    2724 CCCAGGGCGAGGATCCCCCGGTGCA 2748
Db   19736 AGCAGCAGGAGGAGCAGGAGCAGCA 19712

RESULT 15
US-09-335-409-1/c
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match              0.6%;  Score 40.8;  DB 3;  Length 68750;
  Best Local Similarity    46.8%;  Pred. No. 14;
  Matches  163;  Conservative    0;  Mismatches  182;  Indels    3;  Gaps    1;

Qy     569 CTGTGGGCTCAGCGGCGCTCACAGGCCCTCGCTTATTTTGTACCGTGGCCTCCGAGCCTGT 628
Db   50576 CTGCCCTCGCCGCAGCGAGCCCAGCGAGCGCCCTTTCTTCCCCGCCGTCTGGATCTC 50517

Qy     629 ATCACCCGGCAGCGCCTCCGTCTCCTGCCCCCGATGCAGGCTCGTGTGCGTGGGCTCCAG 688
Db   50516 GTCCAGGGCGGCATCAGGATCGGCGCGGGTCATCTCGATGAACAGCGCGGGGCCACC 50457

Qy     689 GCCAGGAAGCGATATCTCCAGCGGAGGTCAGCTCTGGGACAGCTGAACACCATTCTCCTA 748
Db   50456 CTCCAGCAGCGCTTGCGCCGCCGCAGCGAA[illegible] 50397

Qy     749 GTGGCCCGGCCCCTGCTCCGGAGAGGAGAGAAGCTAG[illegible]CAGC 808
Db   50396 GTAGCTCGCATCGAGCTCCGGACC[illegible]GCGGATTACCCCGTCCGTCACCGTCGAGCGCATC-- 50339

Qy     809 GGGGAGCCCTGGGGGAAAGTGTCAAATATGGACCTGGGTCCCTTAGAGATCCCCCCCCAG 868
Db   50338 -GGCACCGCAGCCGCTCGCGGCCGGATCGTC[illegible]GCTTCGCGCAG 50280

Qy     869 CTGGCTACTCTGCTGGAGAGGGCGGAAGGCCACCAGGCCTTGCTGACG 916
Db   50279 CGGGTCGACCTGCGGGCTATGGCTGGCGAGG[illegible] 50232

Search completed: January 10, 2003, 02:13:03
Job time : 829.767 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on January 9, 2003, 06:19:32 ; Search time 242.238 Seconds
(without alignments)
11430.411 Million cell updates/sec

Title: US-09-803-126-2
Perfect score: 6293
Sequence 1 cgctgggactgtcacctacc ... aaaaaaaaaaaaaaaaaaaa 6293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched 388086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6293	100.0	6293	10	US-09-803-126-2	Sequence 2, Appli
2	4350	69.2	4375	10	US-09-803-126-3	Sequence 3, Appli
3	1787.6	28.4	4174	10	US-09-803-126-5	Sequence 5, Appli
4	1490	23.7	3780	10	US-09-803-126-7	Sequence 7, Appli
5	913	14.5	1960	10	US-09-822-849A-265	Sequence 265, App
6	621.2	9.9	1247	9	US-09-764-868-145	Sequence 145, App
c 7	166.2	2.6	24757	9	US-09-764-868-1347	Sequence 1347, Ap
c 8	94.4	1.5	566	10	US-09-864-761-8013	Sequence 8013, Ap
c 9	92.2	1.5	233	10	US-09-864-761-24717	Sequence 24717, A
c 10	83.4	1.4	526	10	US-09-864-761-7087	Sequence 7087, Ap
c 11	75.2	1.2	170	10	US-09-864-761-23818	Sequence 23818, A
12	63.4	1.0	312	10	US-09-783-590-1780	Sequence 1780, Ap
13	62	1.0	3836	12	US-10-002-600-7	Sequence 7, Appli
14	52	0.8	7468	10	US-09-890-107-3357	Sequence 3357, Ap
15	47.4	0.8	145	10	US-09-783-590-3077	Sequence 3077, Ap
16	47	0.7	316	10	US-09-864-761-17473	Sequence 17473, A
17	47	0.7	436	10	US-09-864-761-686	Sequence 686, App
18	44	0.7	5500	9	US-09-902-525-38	Sequence 38, Appl
c 19	43.2	0.7	446	10	US-09-983-965-4231	Sequence 4231, Ap
20	43.2	0.7	2038	10	US-09-880-107-2102	Sequence 2102, Ap
c 21	42.8	0.7	2379	10	US-09-781-080B-12	Sequence 12, Appl
22	42.8	0.7	2802	10	US-09-925-300-238	Sequence 238, App
23	41.8	0.7	349	10	US-09-867-701-9513	Sequence 9513, Ap
c 24	41.8	0.7	461	10	US-09-864-761-14656	Sequence 14656, A
c 25	41.8	0.7	526	10	US-09-864-761-31189	Sequence 31189, A
26	41.6	0.7	1322	10	US-09-925-297-26	Sequence 26, Appl
27	41.4	0.7	4509	9	US-10-124-800-5	Sequence 5, Appli
28	41.4	0.7	4512	9	US-10-124-800-27	Sequence 27, Appl
29	41	0.7	329	10	US-09-960-352-12421	Sequence 12421, A
c 30	41	0.7	393	9	US-10-001-873-15	Sequence 15, Appl
31	41	0.7	440	10	US-09-864-761-3517	Sequence 3517, Ap
c 32	40.8	0.6	68750	9	US-10-014-717-1	Sequence 1, Appli
33	40.6	0.6	6604	10	US-09-880-107-1749	Sequence 1748, Ap
34	40.4	0.6	479	10	US-09-764-869-351	Sequence 350, App
35	40.4	0.6	2773	9	US-09-992-598-178	Sequence 178, App
36	40.4	0.6	2773	9	US-09-989-293A-178	Sequence 178, App
37	40.4	0.6	2773	9	US-10-062-547-33	Sequence 33, Appl
38	40.4	0.6	2773	9	US-09-989-735-178	Sequence 178, App
39	40.4	0.6	2773	9	US-09-990-444-178	Sequence 178, App
40	40.4	0.6	2773	9	US-09-989-730-178	Sequence 178, App
41	40.4	0.6	2773	9	US-09-990-436-178	Sequence 178, App
42	40.4	0.6	2773	9	US-09-991-181-178	Sequence 178, App
43	40.4	0.6	2773	9	US-09-993-687-178	Sequence 178, App
44	40.4	0.6	2773	9	US-09-989-734-178	Sequence 178, App
45	40.4	0.6	2773	10	US-09-989-722-178	Sequence 178, App

ALIGNMENTS

RESULT 1
US-09-803-126-2
; Sequence 2, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 6293
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse myosin related protein (MRP) variant 1 cDNA
US-09-803-126-2

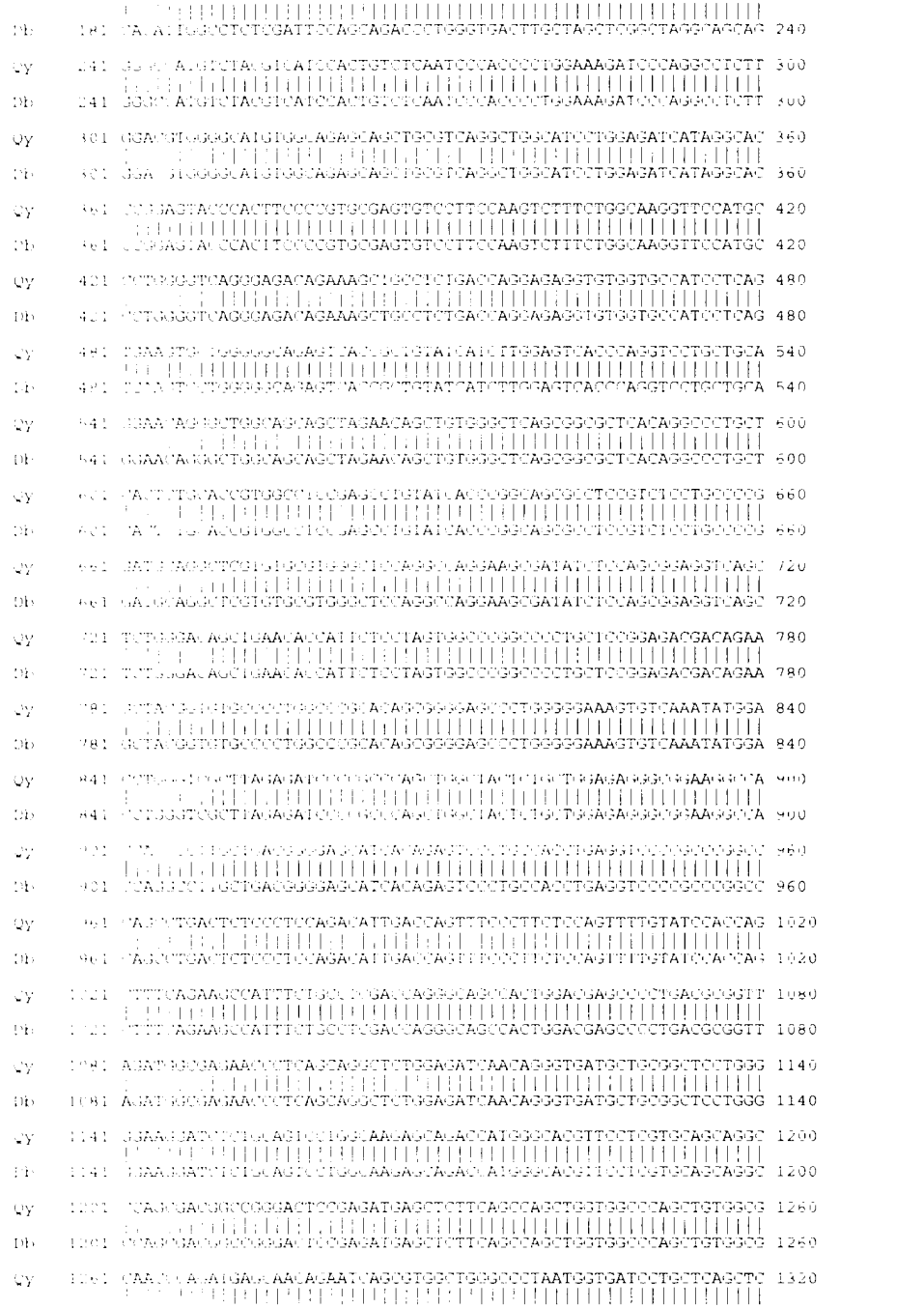
Query Match 100.0%; Score 6293; DB 10; Length 6293;
Best Local Similarity 100.0%; Pred No. 0;
Matches 6293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

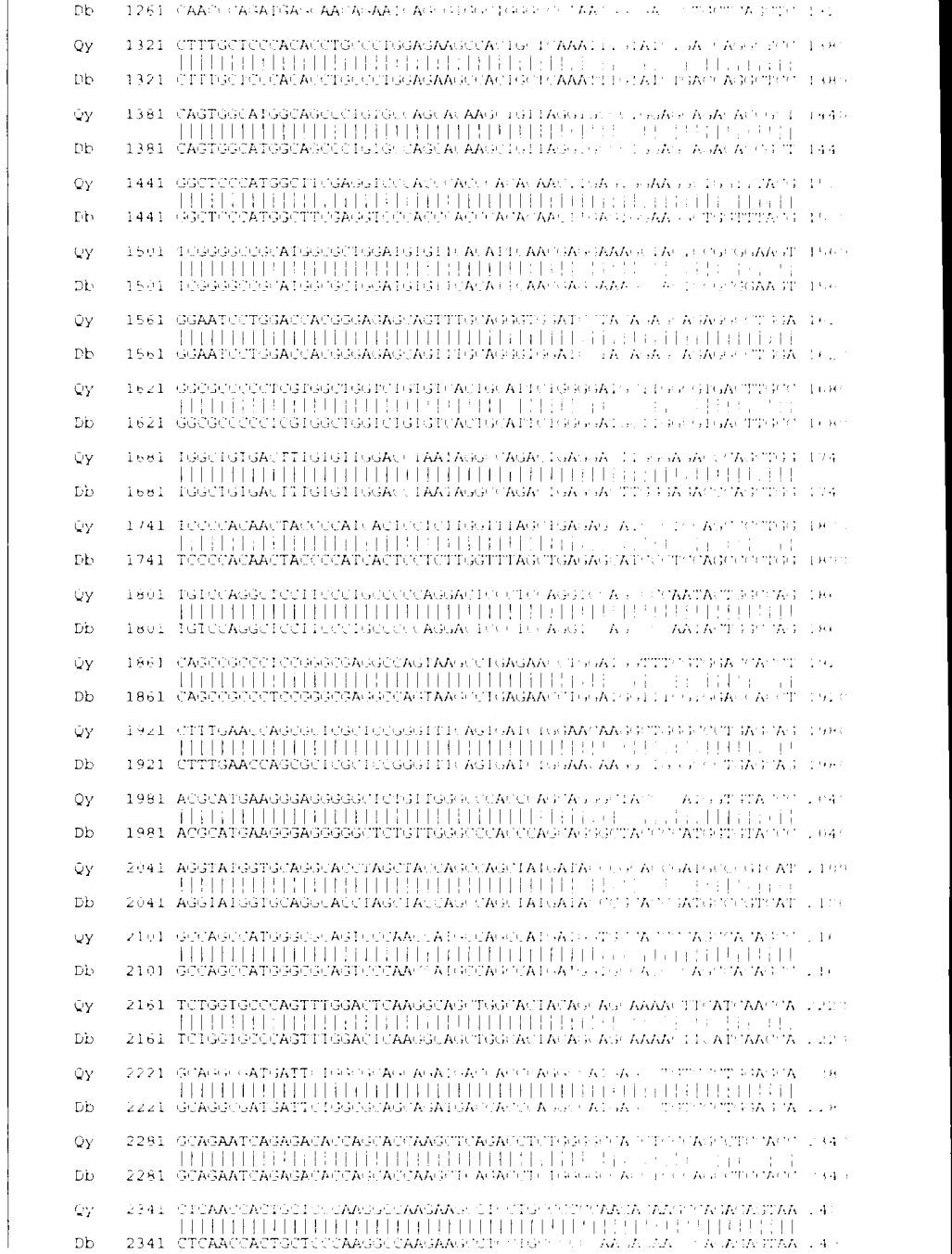
```
Qy       1 CGCTGGGACTGTCACCTACCAGGTGCACAAGTTCATAAACAGAAACAGGGGCCAGCTGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCTGGGACTGTCACCTACCAGGTGCACAAGTTCATAAACAGAAACAGGGGCCAGCTGGA 60

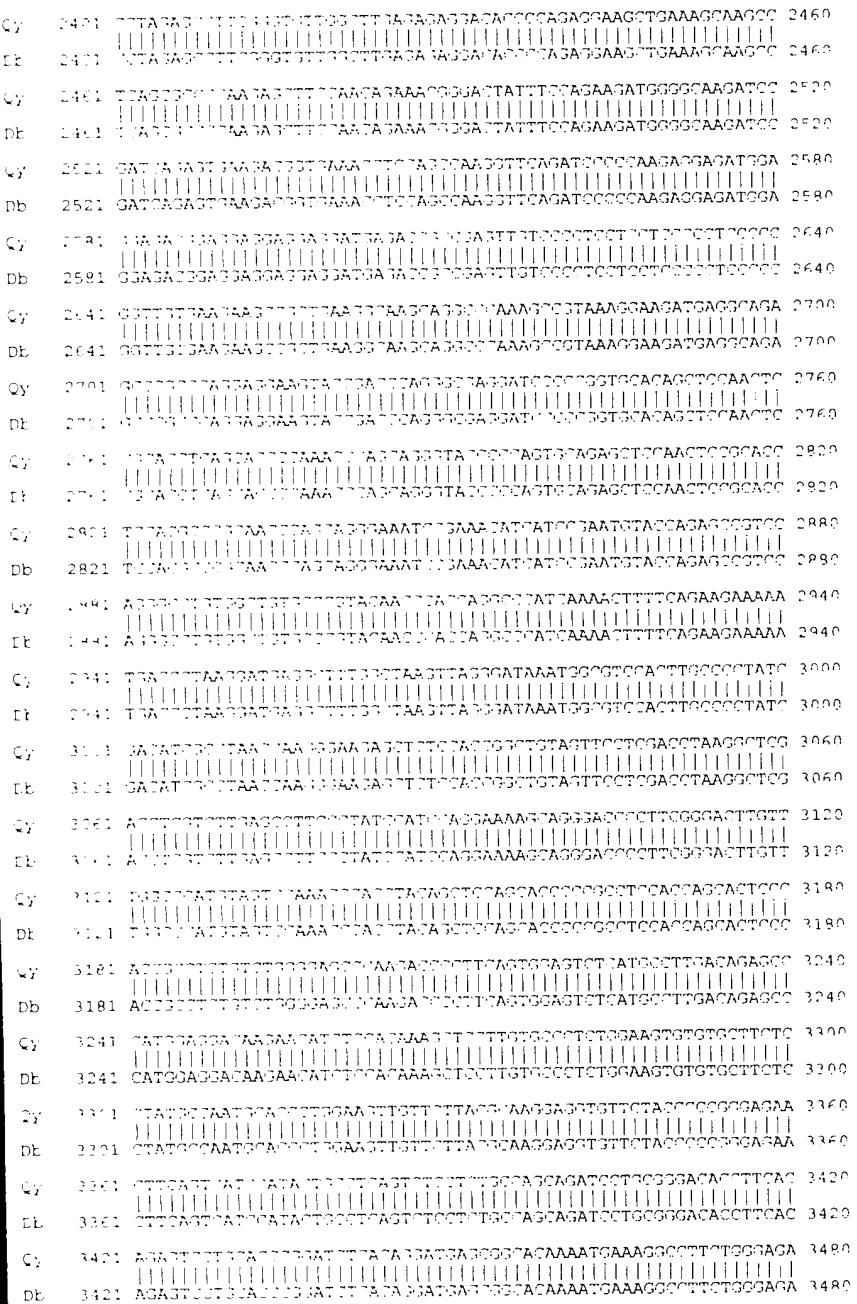
Qy      61 CCCCGCTGTGCTGGAGATGCTCAGGCAGAGCCAGCTTCAGGTGACTTAGCCTTCCTTTCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCCCGCTGTGCTGGAGATGCTCAGGCAGAGCCAGCTTCAGGTGACTTAGCCTTCCTTTCA 120

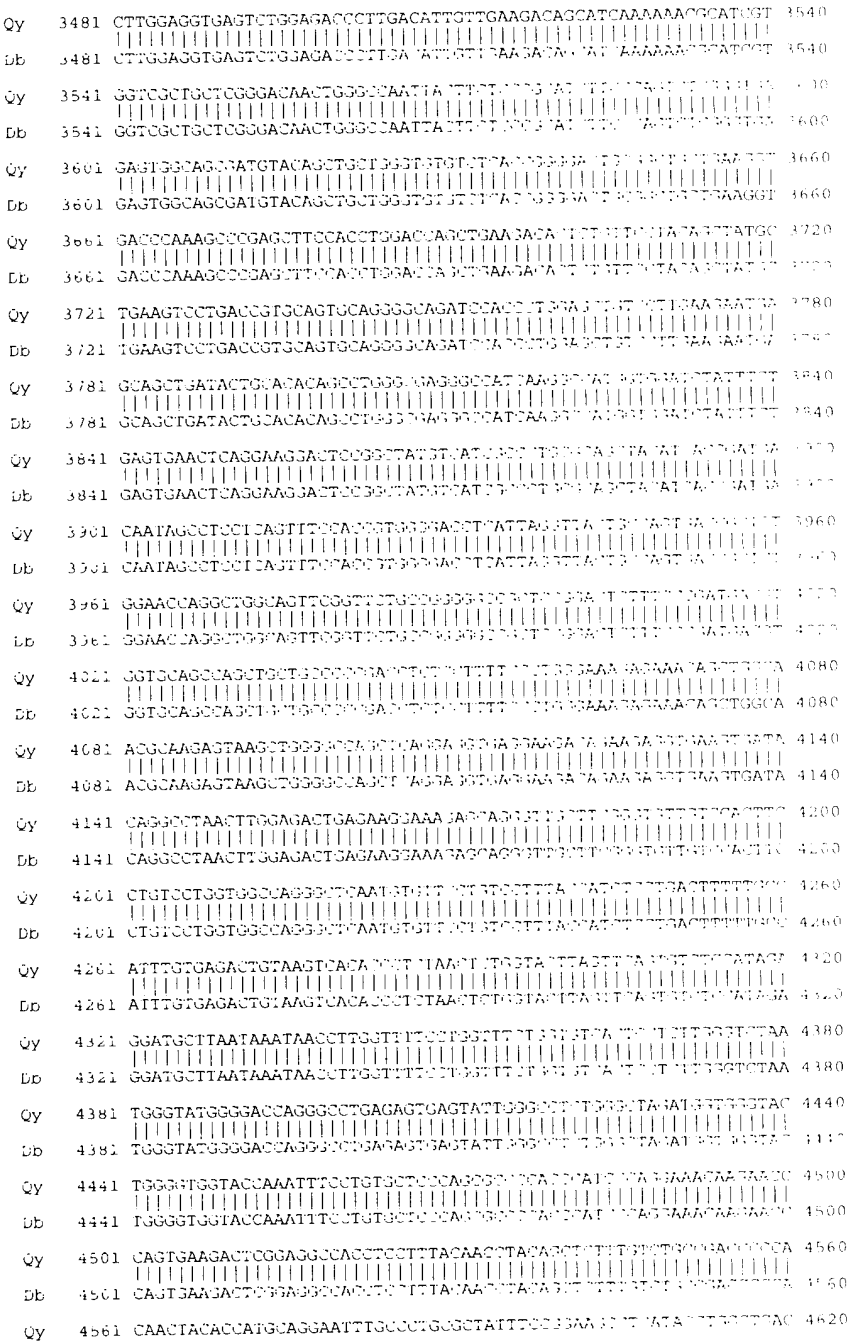
Qy     121 GCTCATGGGCAGCCTGTTCCAAGAAGCAGAGCCCAGCTTGAACTGACAAAACAAACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTCATGGGCAGCCTGTTCCAAGAAGCAGAGCCCAGCTTGAACTGACAAAACAAACC 180

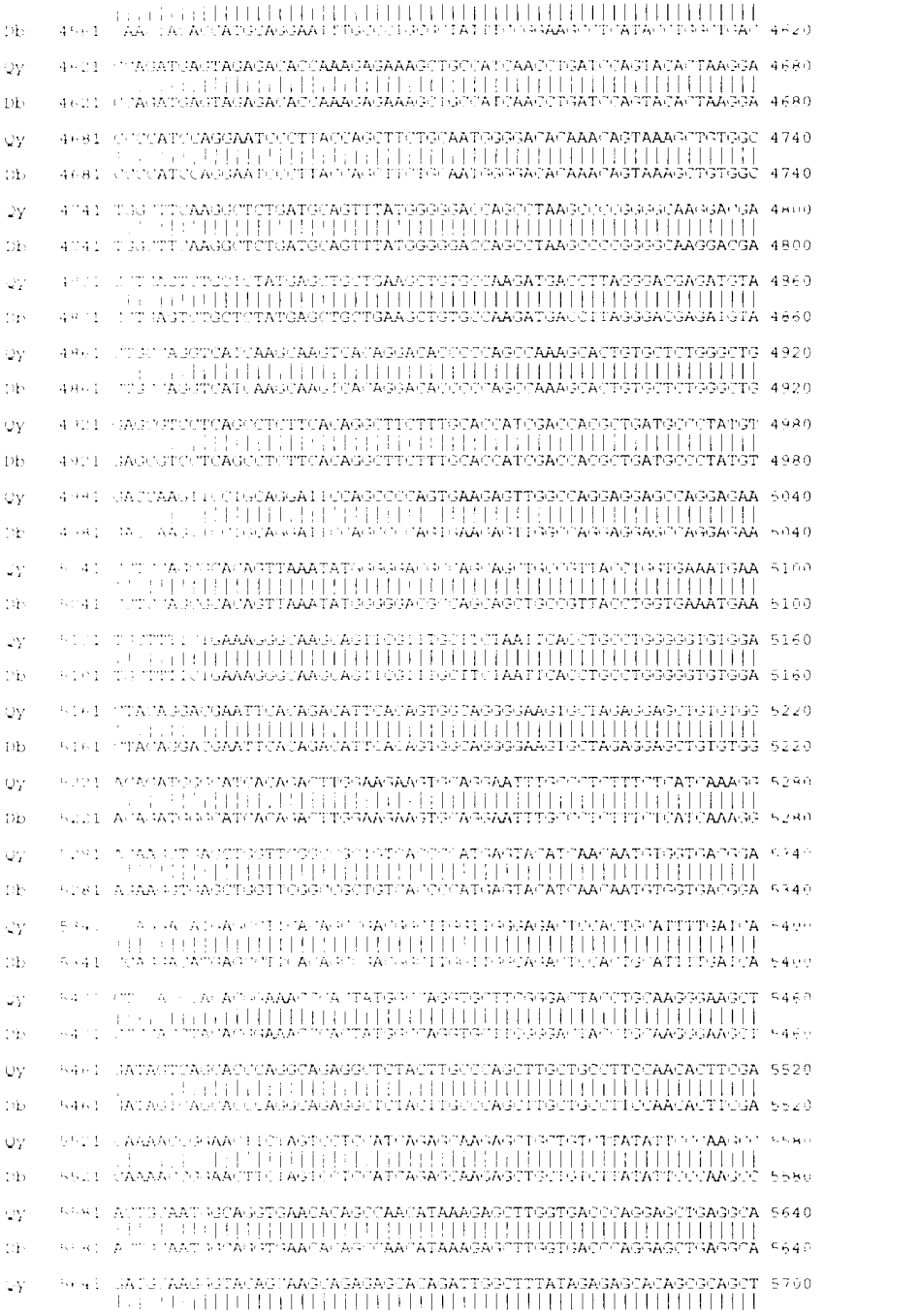
Qy     181 CACATTGGCCTCTCGATTCCAGCAGACCCTGGGTGACTTGCTAGCTTTTTAGCCAGCAG 240
```







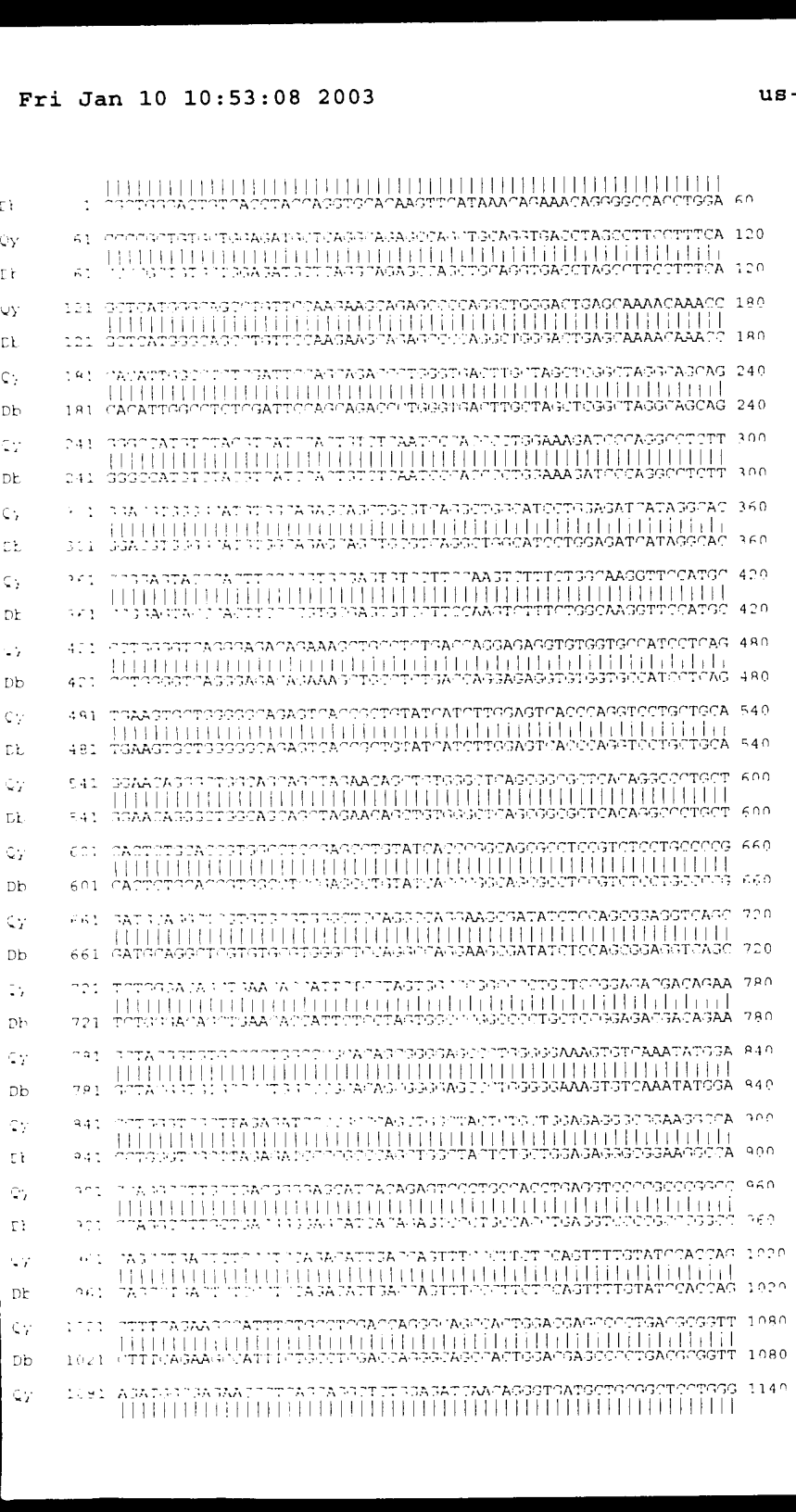


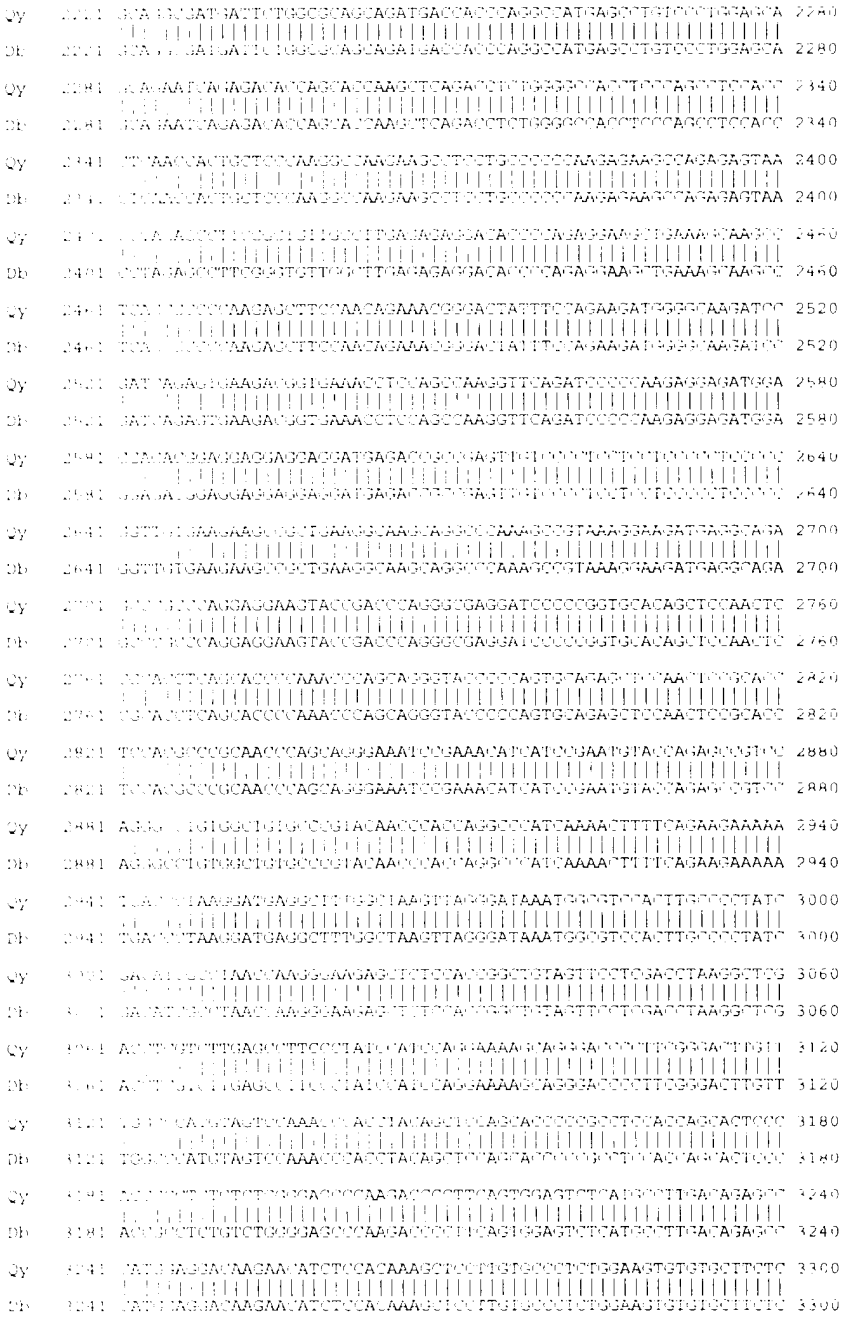


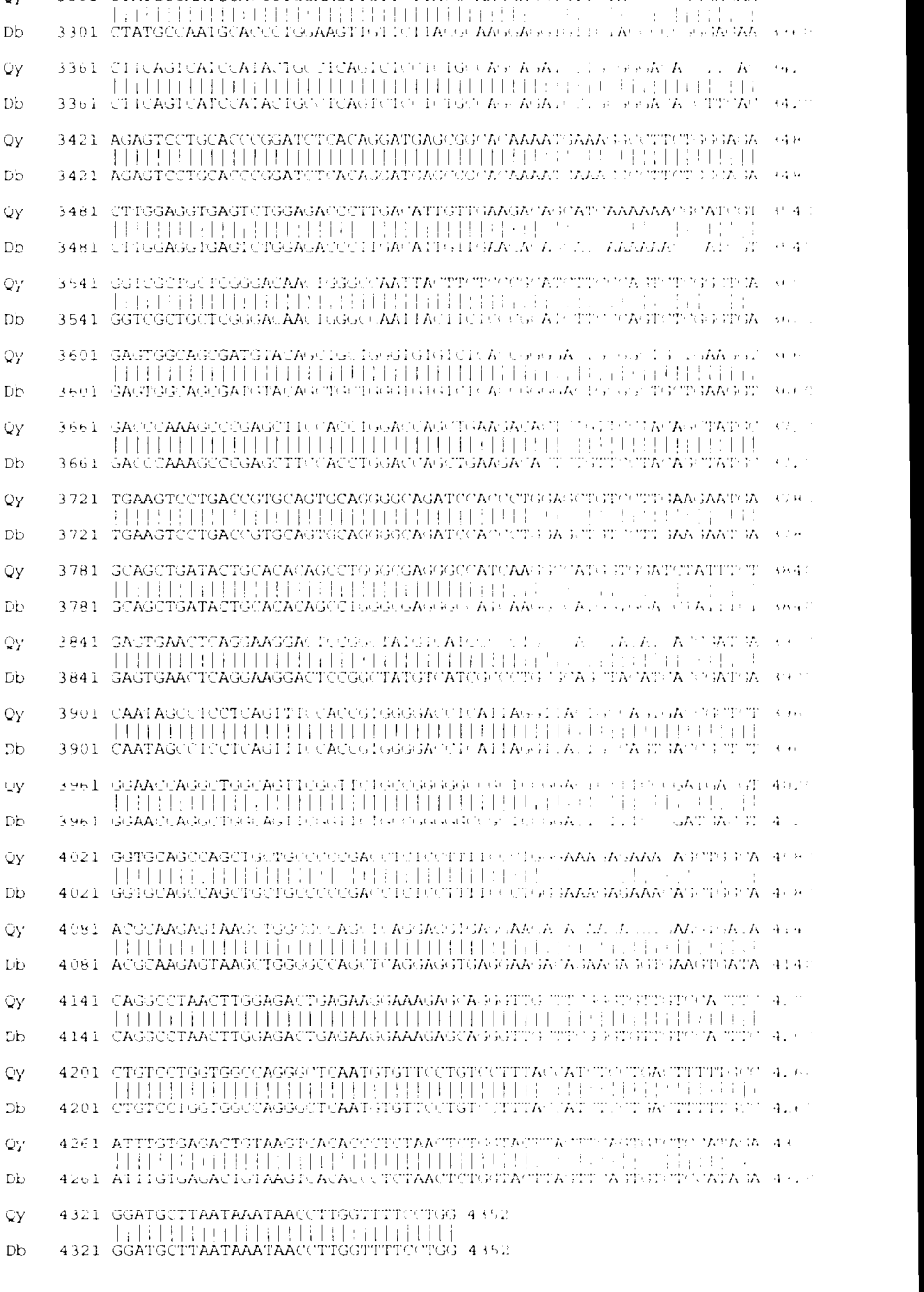
RESULT 2
US-09-803-126-3
; Sequence 3, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Peng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015402-000210US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4375
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse myosin related protein (MRP) variant 3 cDNA
US-09-803-126-3

Query Match 69.2%; Score 4352; DB 1; Length 4375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCTGGGACTGTCACCTACCAGGTGCACAAGTTCATAAACAAAAA 4352

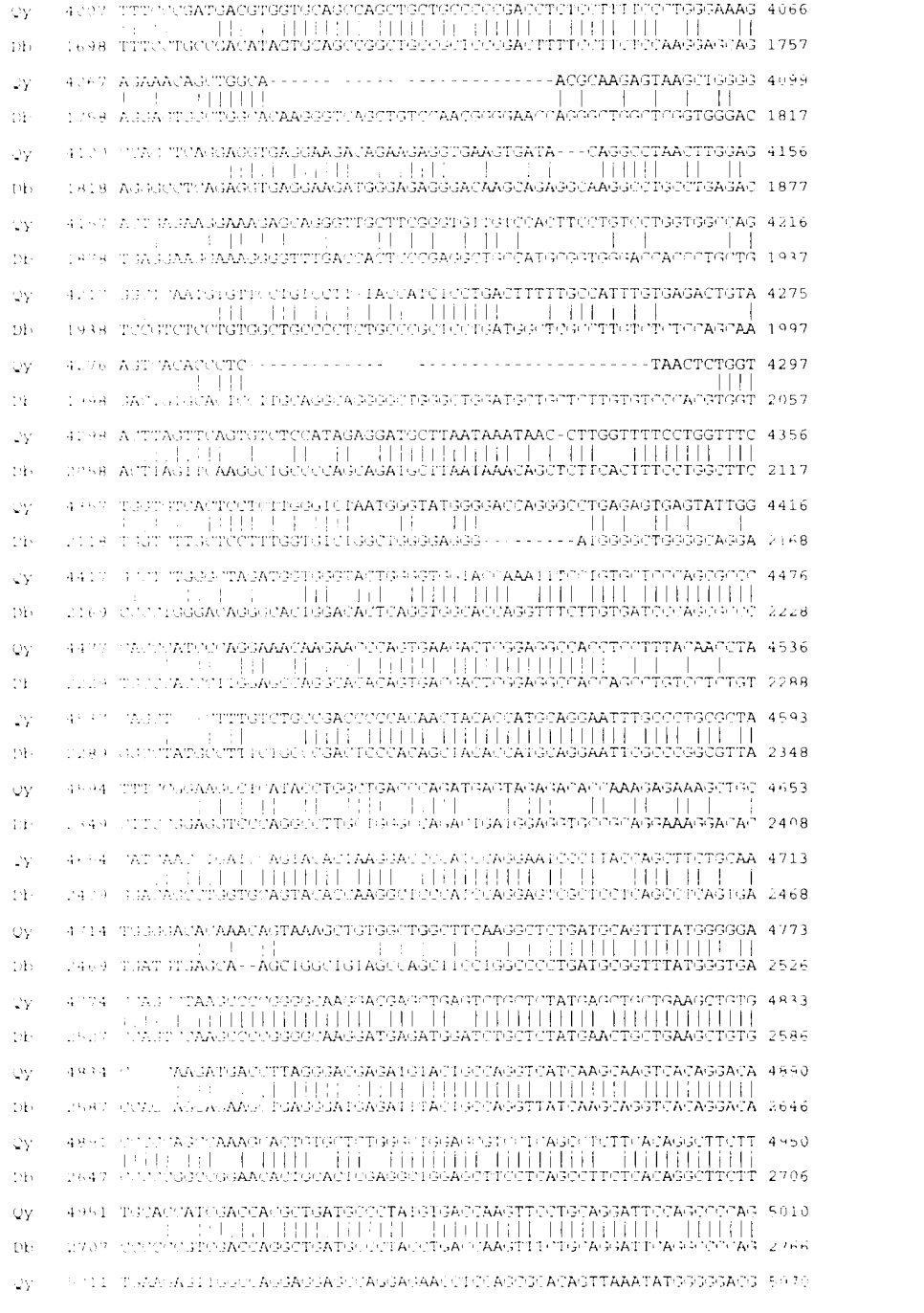






RESULT 3
US-09-803-126-5
; Sequence 5, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION
; APPLICANT: Brooks, Alan R.
; APPLICANT: Penn, Gary G
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015203.000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related
; OTHER INFORMATION: protein variant 1 (hMRP1)
US-09-803-126-5

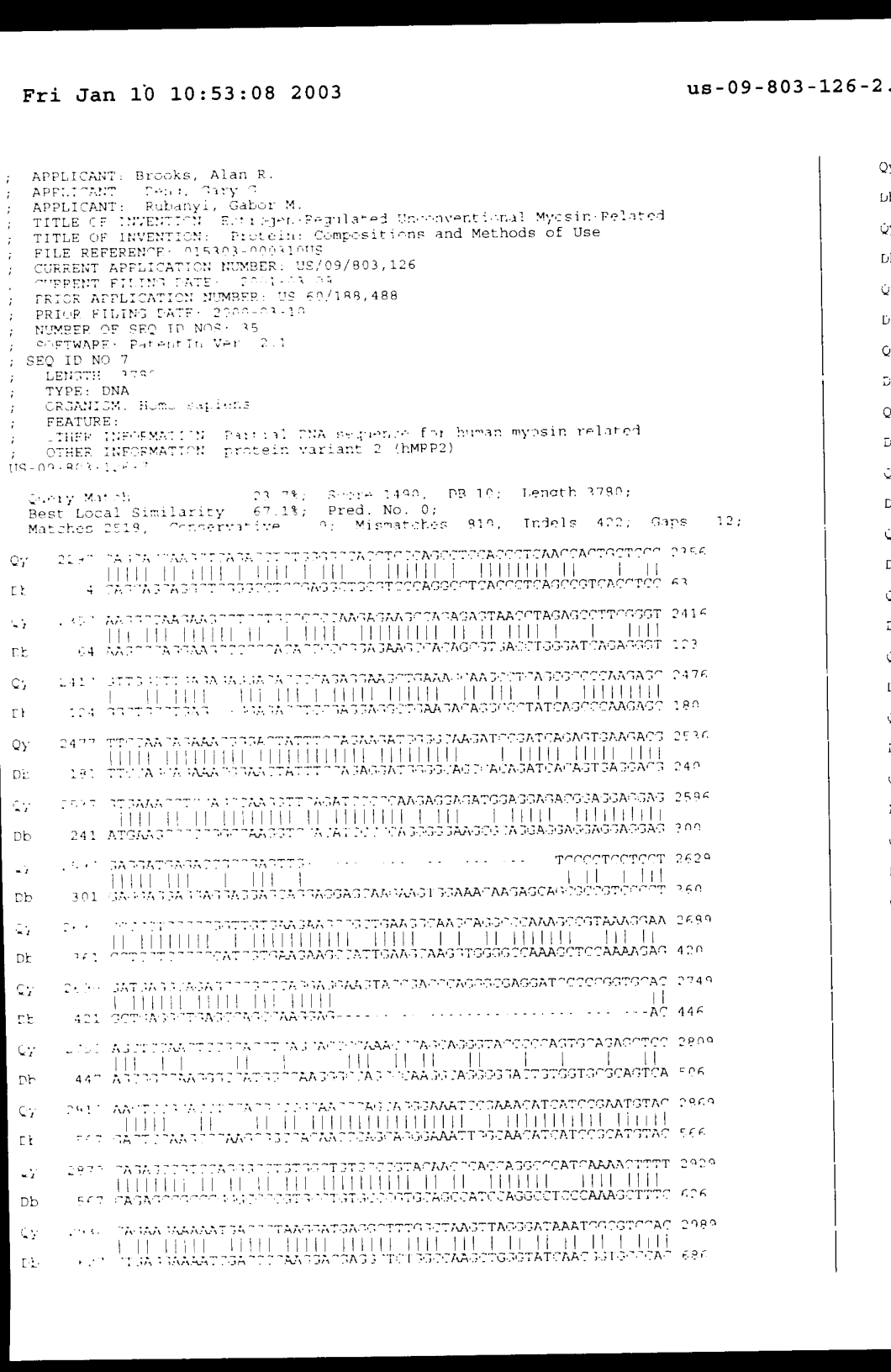
Query Match 29.4%; Score 1297.6; DB 10; Length 4174;
Best Local Similarity 70.3%; Pred. No. 0;
Matches 2685; Conservative 0; Mismatches 969; Indels 166; Gaps 15;

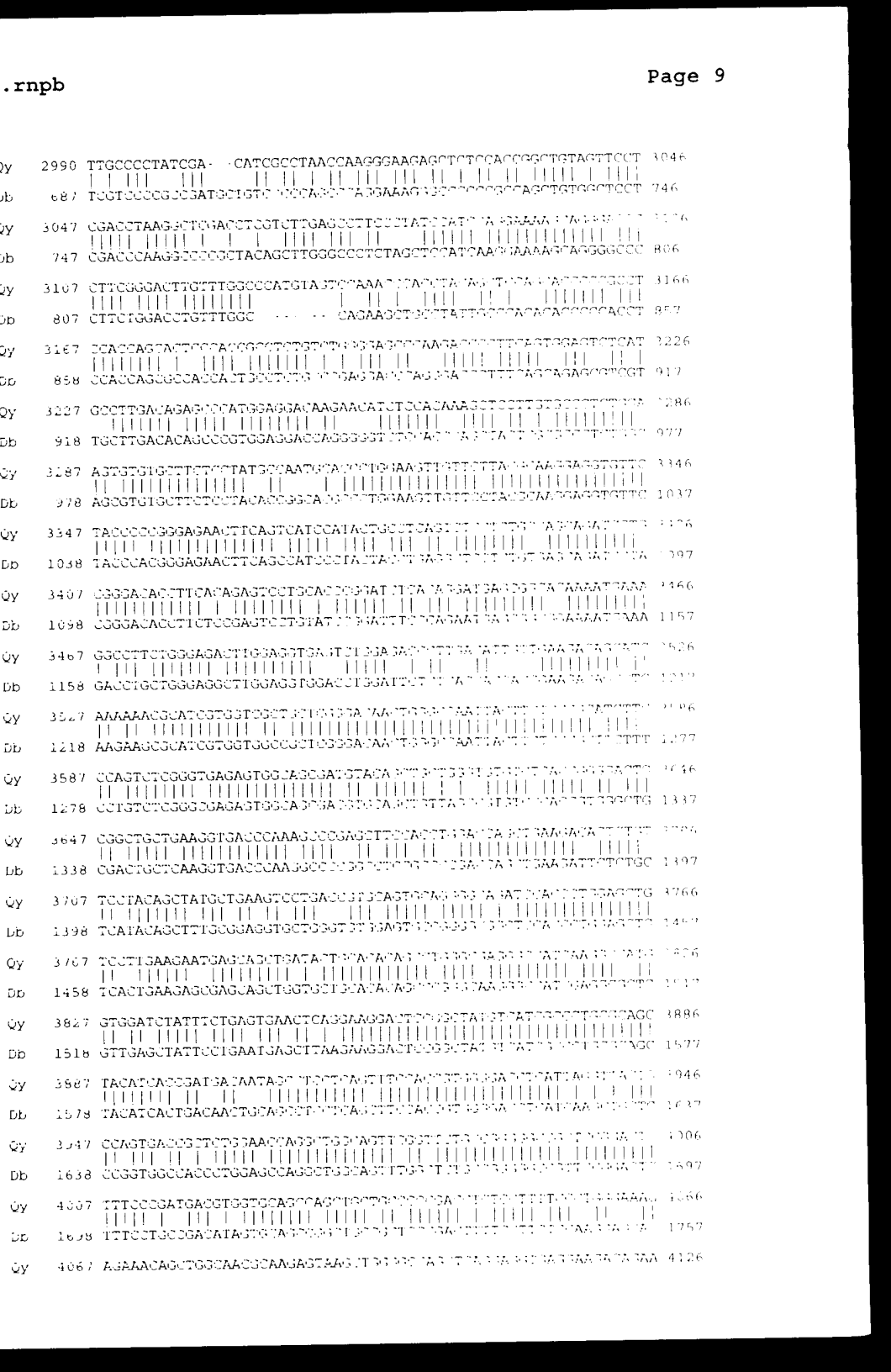


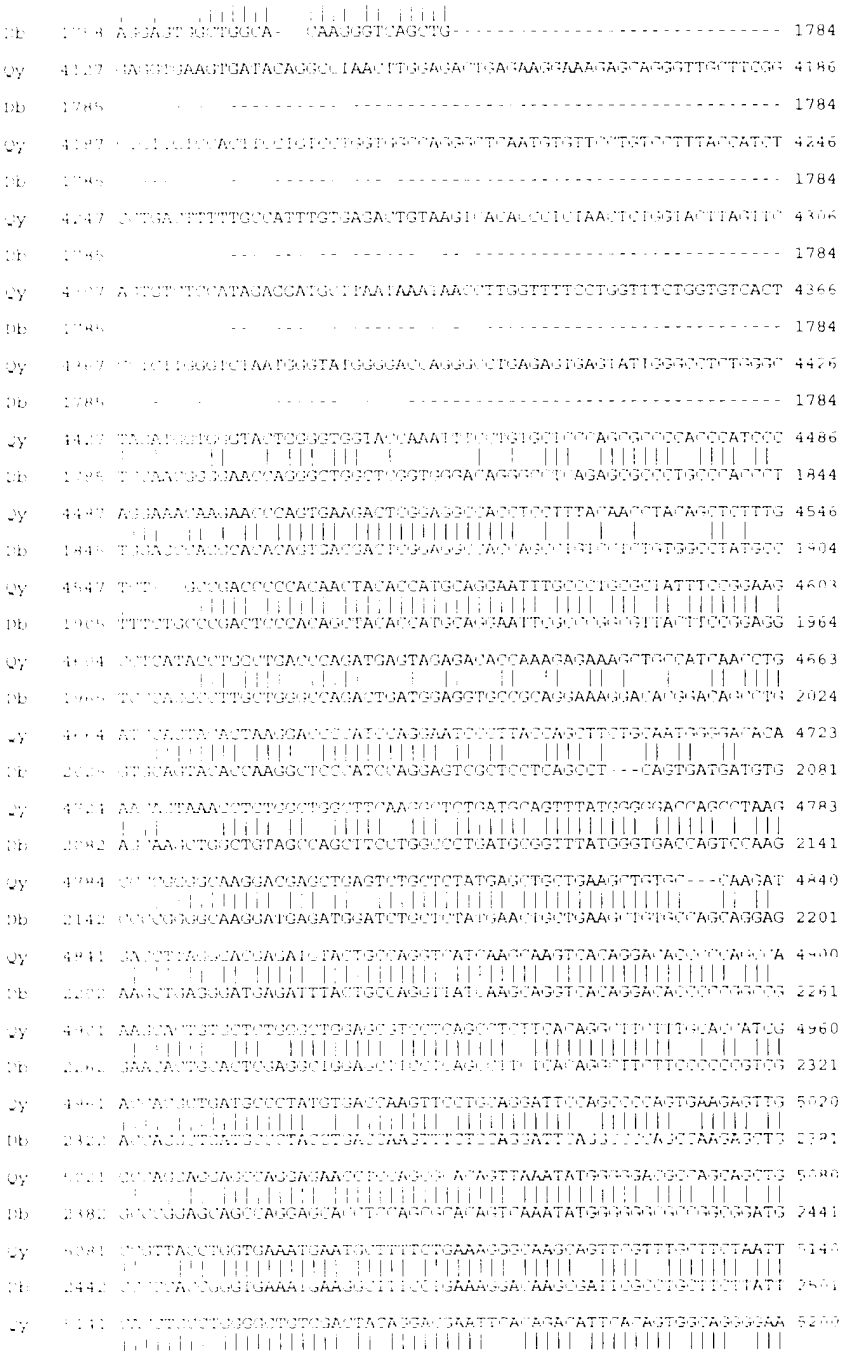
RESULT 4
US-09-803-126-7
; Sequence 7, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:

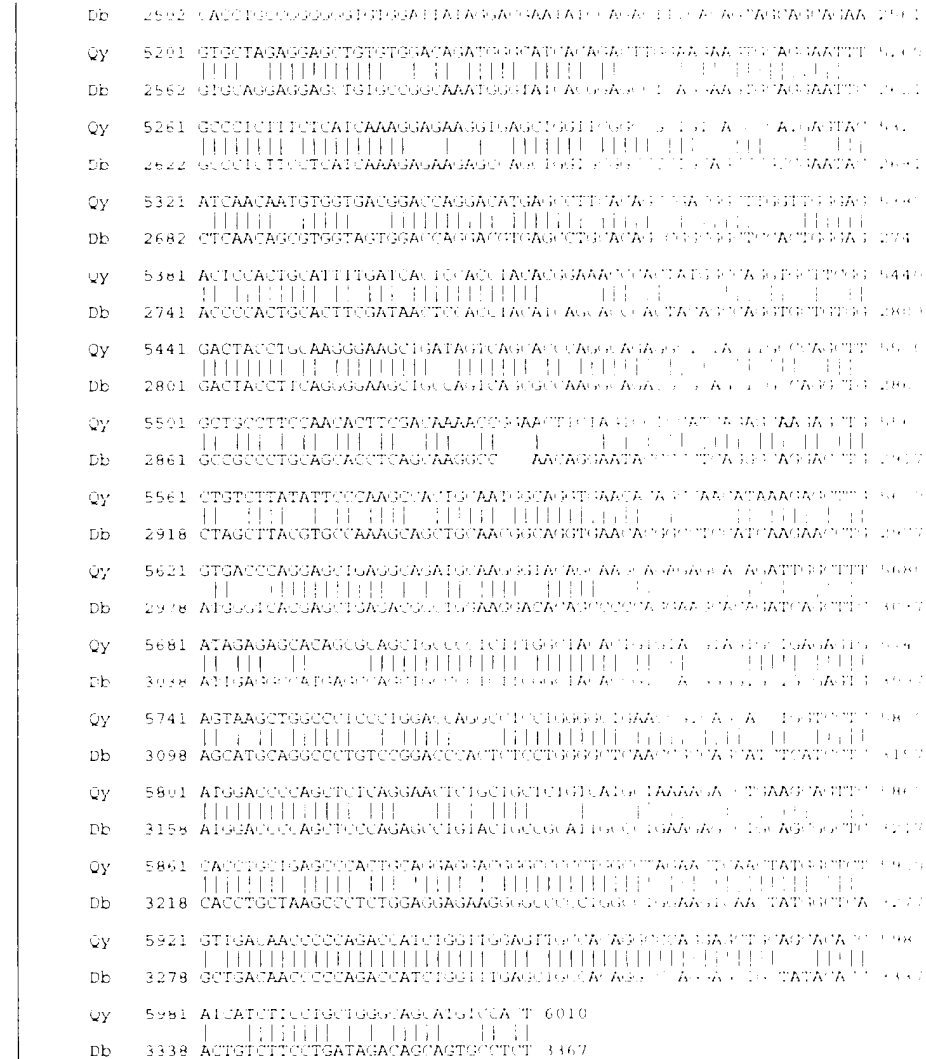
; APPLICANT: Brooks, Alan R.
; APPLICANT: Lewis, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000410US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial DNA sequence for human myosin related
; OTHER INFORMATION: protein variant 2 (hMPP2)
US-09-803-126-2

Query Match 23.7%; Score 1490; DB 10; Length 3780;
Best Local Similarity 67.1%; Pred. No. 0;
Matches 2519; Conservative 0; Mismatches 810; Indels 402; Gaps 12;

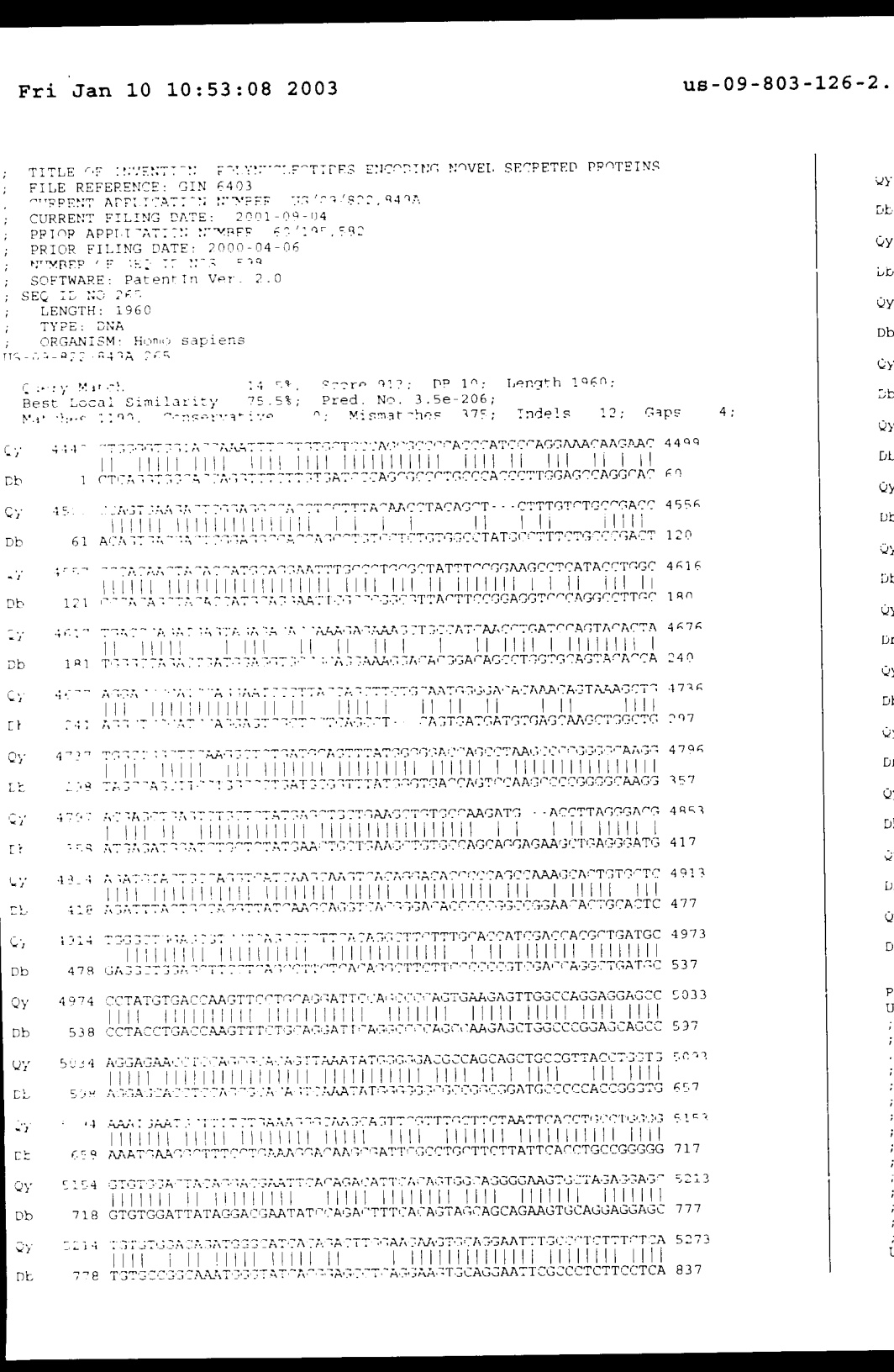


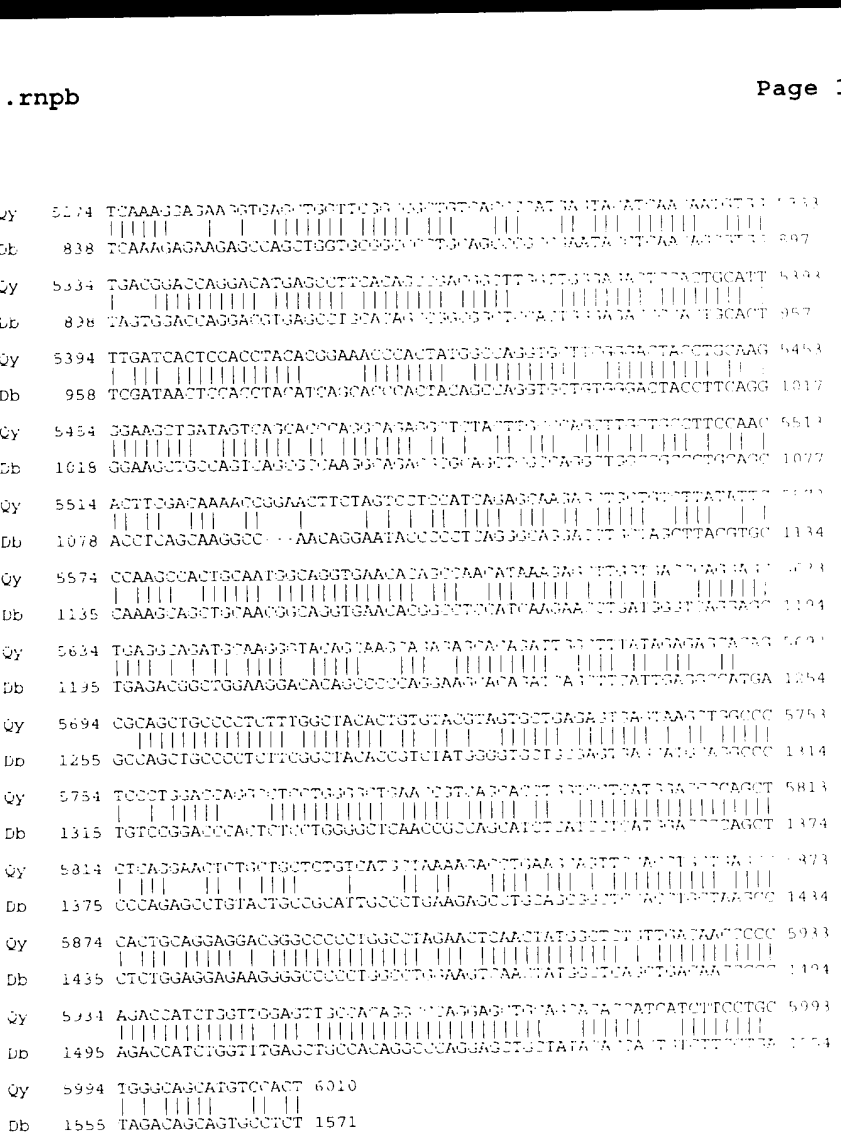






RESULT 5
US-09-822-849A-265
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Kosiewicz, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.


; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/803,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ IDS: 539
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 265
; LENGTH: 1960
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-803-849A-265
Query Match 14.5%; Score 913; DB 10; Length 1960;
Best Local Similarity 75.5%; Pred. No. 3.5e-206;
Matches 1190; Conservative 0; Mismatches 375; Indels 12; Gaps 4;


PESULT 6
US-09-764-868-145
; Sequence 145, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver 2.0
; SEQ ID NO 145
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-145

```
  Query Match             9.9%;  Score 621.2;  DB 9;  Length 1247;
  Best Local Similarity   80.6%;  Pred. No. 2.3e-137;
  Matches  [illegible];  Conservative    1;  Mismatches  174;  Indels    0;  Gaps    0;

Qy    [illegible]

Db       1 [illegible]

Qy    [illegible]

Db      61 [illegible]

Qy    [illegible]

Db     121 [illegible]

Qy    [illegible]

Db     181 [illegible]

Qy    [illegible]

Db     241 [illegible]

Qy    [illegible]

Db     301 [illegible]

Qy    [illegible]

Db     361 [illegible]

Qy    [illegible]

Db     421 [illegible]

Qy    [illegible]

Db     481 [illegible]

Qy    [illegible]

Db     541 [illegible]

Qy    [illegible]

Db     601 [illegible]

Qy    [illegible]

Db     661 [illegible]

Qy    [illegible]

Db     721 [illegible]

Qy    [illegible]

Db     781 [illegible]

Qy    [illegible]

Db     841 [illegible]

RESULT 7
[illegible]
; Sequence 1347, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Roach et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PT292
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed  Refer to PALM [illegible] wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1347
;  LENGTH: 24757
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-764-868-1347

  Query Match             2.6%;  Score 166.2;  DB 9;  Length 24757;
  Best Local Similarity   80.2%;  Pred. No. 5e-29;
  Matches  195;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy     1119 GGGTGATGCTGCGGCTCCTGGGGCAAGGATCTCTGTAGTCTTGTAAGAGTAGACTATGG 1178
            ||  |||||||||||||||||||| |||| ||| |||||| | |||||| |||
Db    17652 GGCAGATGCTGCGGCTCCTGGGGCACGGATCTCTGTAGTCTTGTAAGAGTAGACTATGG 17593

Qy     1179 GCACGTTCGTCGTGCAGCAGGCCAGGCAGCAGTTGAC[illegible]
            || | | ||| |||| |||||  ||| | |||| || | |  | |||||||| ||||
Db    17592 GCGCATACCTGGTGCGGCAGGGGCAGTGCCGGCTAGGCTGTTGAATGATTTTTTATTT 17533

Qy     1239 AGCTGCTGGGCCCAGCTGTGGGCAACCCAGATGAGAACAGCAGTTAAGGTGGTGGTTGGTTT 1298
            |||||||||||||||| |||  |||||||||| || | |  |||||||||||||||||||
Db    17532 AGCTGGTGGCCCAGCTATGGCAGAACCCAGATGAACAGCAGATTAAGGTGGTGGTTGGTTT 17473

Qy     1299 TAATGGTGATGCTGCTAGTTGCTTTGCTCCACAGTGGC TTTAAAAGTACAGCCTA 1358
            | ||||   | |||||||| |||| | ||   |||| || | |||||||||||||
Db    17472 TCATGGCTGTTTTGCTCAGCGCCTTTCCCCACTGGCTTT CTAAAAGTACAGCCTA 17413

Qy     1359 AAT 1361
            |||
Db    17412 AGT 17410

RESULT 8
US-09-864-761-8013/c
; Sequence 8013, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Funk, David E.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID [illegible]
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MASS ARRAY
;  FILE REFERENCE: Aeomica X 1
;  CURRENT APPLICATION NUMBER:  US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER:  US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER:  US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER:  US 60/223,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER:  US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER:  PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER:  PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8013
; LENGTH: 566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
US-09-864-761-8013

Query Match 1.5%; Score 94.4; DB 10; Length 566;
Best Local Similarity 76.2%; Pred. No. 7.6e-13;
Matches 115; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 5282 GAATTTGATTTGTTTGTTTGTTTCATGCCCATGAGTACATCAACAATGTGGTGACGGAC 5341
Db 381 GAATACGTCAACAGCGTGGTAGTGGAC 324

Qy 5342 CAGGATATGATTTTTATATTTGATGTTTGTTTGGAGACTGCACTGCATTTGATCAC 5401
Db 323 CAGGACGTGATGTTTGTATGCATGTTGTCACTGGGAGACCCCACTGCACTTCGATAAC 264

Qy 5402 TCCACCTACACGGAAACCCACTATGGCCAGGT 5433
Db 263 TCCACCTACATCACATACATCAGGT 232

RESULT 9
US-09-864-761-24717/c
; Sequence 24717, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24717
; LENGTH: 233
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019214.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: SWISSPROT HIT P70085, EVALUE 2.90e-01
; OTHER INFORMATION: NT HIT: AE004636.1, EVALUE 8.70e+00
; OTHER INFORMATION: EST_HUMAN HIT AW752407.1, EVALUE 3.50e-78
US-09-864-761-24717

Query Match 1.5%; Score 92.2; DB 10; Length 233;
Best Local Similarity 77.2%; Pred. No. 1.6e-12;
Matches 112; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 5289 AGCTGGTTCGGCCGCTGTCACCCCATGAGTACATCAACAATGTGGTGACGGACCAGGACA 5348
Db 230 AGCTGGTCGGGCCGCCTGCAGCCGGCCGAATACGTCAACAGCGTGGTAGTGGACCAGGACG 171

Qy 5349 TGAGCCTTCACAGCCGGAGGCTTGGTTGGGAGACTGCACTGCATTTGATCACTCCACCT 5408
Db 170 TGAGCCTGCACAGCCGGCGGCTCCACTGGGAGACCCCACTGCACTTCGATAACTCCACCT 111

Qy 5409 ACACGGAAACCCACTATGGCCAGGT 5433
Db 110 ACATCAGCACCCACTACAGCCAGGT 86

RESULT 10
US-09-864-761-7087/c
; Sequence 7087, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1

```
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: AceMax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23817
;   LENGTH: 224
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
US-09-864-761-23817

  Query Match           1.1%;  Score 89.4;  DB 10;  Length 224;
  Best Local Similarity   86.1%;  Pred. No. 1.1e-11;
  Matches   99;  Conservative    9;  Mismatches   16;  Indels    0;  Gaps    0;

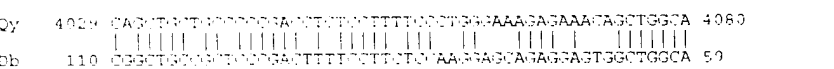
Qy      2814 AGCAATCCTCGCTATCTCATCGTCCTTGCGCAGCTACATCACGCAATGACAATATACCTCTCTA 2877
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       415 AGCAATCCTCGCTATCTCATCGTCCTTGCGCAGCTACATCACGCAATGACAATATACCTCTCTA 356

Qy      2815 GTTTCAGCGTGGGGACCTCATTAGGTTACTGCCAGTGACCGCTCTCTGAAACCAGG 2988
             |||||||||||||||||||||| | | | ||||| ||| || | ||||| |||||
Db       355 GTTTCAACGTGGGGACCTCATCAAGCTACTGCCAGTGACCGCTCTCTGAAACCAGG 301


RESULT 11
US-09-864-761-23819/c
; Sequence 23818, Application US/09864761
; Patent No. US20030048703A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES COPRISING
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: AceMax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23818
;   LENGTH: 170
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019214.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;   OTHER INFORMATION: EST_HUMAN HIT: AV649878.1, EVALUE 5.00e-91
;   OTHER INFORMATION: SWISSPROT HIT: P05805, EVALUE 1.00e-77
;   OTHER INFORMATION: NT HIT: gi11426247, EVALUE 1.50e-80
US-09-864-761-23818

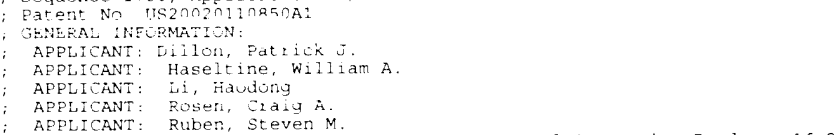
  Query Match           1.7%;  Score 95.2;  DB 1;  Length 170;
  Best Local Similarity   79.5%;  Pred. No. 1.5e-08;
  Matches   89;  Conservative    0;  Mismatches   23;  Indels    4;  Gaps    4;

Qy      2969 GCTGGCAGTTGGCTCTGCCGGGGCCCTCTTCGGACTCTTTCTCCTCATACTATTCAATCTTCA 3032
             ||||||||||| || |||||||||||||||||||||||||||||||| | ||| | |||||||
Db       170 GCTGGCAGTTGGCTCTGCCGGGGCCCTCTTCGGACTCTTTCTCCTCATACTATTCAATCTTCA 111
```



RESULT 12
US-09-783-590-1780
; Sequence 1780, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US/08/[illegible]
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1780
; LENGTH: 312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (26)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (123)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (141)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (142)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (159)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (165)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (169)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (170)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (181)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (196)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (228)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (244)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (290)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1780

Query Match 1.0%; Score 63.4; DB 10; Length 312;
Best Local Similarity 64.3%; Pred. No. 1.2e-05;
Matches 175; Conservative 0; Mismatches 88; Indels 9; Gaps 6;



RESULT 13
US-10-002-600-7
; Sequence 7, Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 3836
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 1095159.4
US-10-002-600-7

Query Match 1.0%; Score 62; DB 12; Length 3836;
Best Local Similarity 63.3%; Pred. No. 8.9e-05;
Matches 95; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy    4043 CTTTAGTGACGAGATGTACTGCCAGGTCATCAAGCAAGTCACAGGACACGCCCCAGCCAAA 4902
           |  |||| |||| |||||| || | |  ||  | | |||
Db     541 CGTCAGCGATGAGATTTACTGCCAGATCTGCAAGCAGCTCTCGGAGAACTTCAAAACAAG 600

Qy    4903 CAATGTTGCTCTGGGCTGGAGCGTCCTCAGCCTCTTCACAGGCTTCTTTGCACCATCGAC 4962
           |  || | |||||||| | | |||||||||| |  |||| ||| ||||  ||
Db     601 CAGCCTGGCCCGGGGCTGGATCCTGCTCAGCCTCTGCCTCGGCTGCTTCCCACCCTCAGA 660

Qy    4963 CACGCTGATGCCCTATGTGACCAAGTTCCT 4992
           | | | |||   ||| |    || ||| |
Db     661 GAGTTCATGAAGTATCTACTGAACTTCAT 690


RESULT 14
US-09-880-107-3867
; Sequence 3867, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Weekley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921 5028 WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,[illegible]
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3970
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3867
;  LENGTH: 7465
;  TYPE: DNA
;  ORGANISM: Homo Sapiens
;  FEATURE:
;  OTHER INFORMATION: Genbank Accession No. US20020142981A1 110226
US-09-880-107-3867

  Query Match             0.8%;  Score 52;  DB 10;  Length 7465;
  Best Local Similarity   56.4%;  Pred. No. 0.028;
  Matches   57;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    4844 CTTAGGGACGAGATGTACTGCCAGGTCATCAAGCAAGTCACAGGACACGCCCCAGCCAAAG 4903
           | |||||| |||| | ||| ||| |  ||  |||||      |
Db    4771 TCCCGGGACGAGATCTACTGCCAGATCAGCAAGCAGCTGACCCACAACCCCTCCAAGAGC 4830

Qy    4904 CAATGTTGCTCTGGGCTGGAGCGTCCTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACC 4963
           |  | | | |||||||  || |    |||| |   |||| || || |
Db    4831 ACTGTCGCCCGGGCTGGATTCTGGTGTCTCTGCGTGGGCTGTTTCGCCCCTCCGAG 4890

Qy    4964 ACGCTGATGCCCTATGTGACCAAGTTCCTGCAGGATCCAGCCCCAGTGAAG 5015
           | |  |  ||  || |  || ||  | || |    |  ||| |  | |
Db    4891 AAGTTTGTCAAGTACCTGCGGAACTTCATCCACGGGGGCCCGCCCCGCTACG 4942


RESULT 15
US-09-783-590-3077
; Sequence 3077, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Baseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,781
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3077
;  LENGTH: 145
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc_feature
;  LOCATION: (4)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc_feature
;  LOCATION: (29)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc_feature
;  LOCATION: (57)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc_feature
;  LOCATION: (138)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc_feature
;  LOCATION: (139)
;  OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3077

  Query Match             0.8%;  Score 47.4;  DB 1;  Length 145;
  Best Local Similarity   69.2%;  Pred. No. 0.051;
  Matches   63;  Conservative    0;  Mismatches   28;  Indels    7;  Gaps    7;

Qy    3990 GGGGGGGCTCCGGAACTCTTCCCCAAGAGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 4049
           |||| |  | | |  |||||| |  ||| | |||||||| |  || || |||| | |
Db      15 GGGGCCCGTTCCGTNTCTTTCCTTCCGACATAGTCCAGCCCCCCCCCCCCCCCCCTTTT 74

Qy    4050 CCTTTTCCCTGCAAAGAGAAACAGCTGGCA 4080
           |||| |||  ||   |||| |   ||| |||
Db      75 CCTTCTCCAAGGAGCAGAGGAGTCGCTGGCA 105


Search completed: January 10, 2003, 02:17:17
Job time : 416.298 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:03:47 ; Search time 7511.93 Seconds
(without alignments)
13567.518 Million cell updates/sec

Title: US-09-803-126-2
Perfect score: 6293
Sequence: 1 [illegible] 875

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16184088 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32369132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	838	13.3	875	14	BQ941051	BQ941051 AGENCOURT
2	828.6	13.2	923	14	BQ956737	BQ956737 AGENCOURT
3	790.2	12.6	981	13	BG968181	BG968181 602835662
4	774.2	12.3	946	14	BQ960468	BQ960468 AGENCOURT
5	721.8	11.5	877	13	BG962322	BG962322 602827123
6	682	10.8	947	14	BQ941605	BQ941605 AGENCOURT
7	670.2	10.6	1006	12	BF582426	BF582426 602094374
8	658	10.5	932	12	BF584026	BF584026 602096267
9	657.6	10.4	677	10	BB072116	BB072116 BB072116
10	657	10.4	669	13	BG962743	BG962743 602830266
11	655.2	10.4	825	13	BG963907	BG963907 602828615
12	653.8	10.4	940	14	BQ924457	BQ924457 AGENCOURT
13	634.8	10.1	638	10	BB559556	BB559556 BB559556
14	620.8	9.9	942	12	BF577925	BF577925 602091945
15	594.2	9.4	676	12	BF581819	BF581819 602090822
16	588.6	9.4	839	13	BG964041	BG964041 602828784
c 17	575	9.1	610	13	BI851062	BI851062 imageqc_1
18	556	8.8	958	12	BF582477	BF582477 602094438
19	552.2	8.8	653	10	BB624737	BB624737 BB624737
20	539.8	8.6	882	12	BF579361	BF579361 602095651
21	536	8.5	929	12	BF585148	BF585148 602101892
22	529	8.4	544	9	AA475108	AA475108 vq94f08.r
23	510.2	8.1	548	9	AA538534	AA538534 vj01d03.r
24	504.4	8.0	1088	14	BM906425	BM906425 AGENCOURT
25	487.6	7.7	720	13	BG964584	BG964584 602832263
26	484	7.7	505	9	AI891960	AI891960 wi57a11.y
c 27	459.6	7.3	698	13	BI294877	BI294877 UI_R_PK0
28	459.6	7.3	920	12	BG680906	BG680906 602628729
29	445.2	7.1	473	10	BB745593	BB745593 BB745593
30	430.2	6.8	603	14	BM791149	BM791149 K_EST0071
31	424.6	6.7	1015	14	BM923763	BM923763 AGENCOURT
32	423.2	6.7	440	10	BB748464	BB748464 BB748464
33	420.8	6.7	728	14	BM720779	BM720779 UI_R_EO1
34	420	6.7	431	10	BB738897	BB738897 BB738897
35	416.8	6.6	642	10	BB666962	BB666962 BB666962
36	413.6	6.6	969	14	BM804816	BM804816 AGENCOURT
37	412.2	6.6	852	12	BF583262	BF583262 602101725
38	411.4	6.5	413	10	BB738674	BB738674 BB738674
39	406.4	6.5	616	13	BG964814	BG964814 602829369
40	405	6.4	406	10	BB738401	BB738401 BB738401
41	401	6.4	422	14	BQ938910	BQ938910 AGENCOURT
42	400.8	6.4	1070	13	BI766913	BI766913 603053164
43	392.4	6.3	407	10	BB739631	BB739631 BB739631
44	393	6.2	908	13	BI766411	BI766411 603052657
c 45	386.8	6.1	581	10	BE116910	BE116910 UI_R_BS1

ALIGNMENTS

RESULT 1
BQ941051
LOCUS BQ941051 875 bp mRNA linear EST 21 AUG 2002
DEFINITION AGENCOURT_8882063 NCI_CGAP_Ov24 Mus musculus cDNA clone IMAGE:6476207 5', mRNA sequence.
ACCESSION BQ941051
VERSION BQ941051.1 GI:22356529
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 875)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14016 row: c column: 24
High quality sequence stop: 634.
FEATURES Location/Qualifiers
source 1..875

```
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6476207"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally  Primer: Oligo dT
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      165 a    276 c    261 g    172 t      1 others
ORIGIN

  Query Match             13.3%;  Score 838;  DB 14;  Length 875;
  Best Local Similarity   98.1%;  Pred. No. 2.4e-131;
  Matches  838;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy    1128 ...GAGGGGTTAGATGGGAGAACCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCT 1188
           ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCTGACTGGTTAGATGGGAGAACCTTCAGCAGGCTCTGGAGATCAACAGGGTGATGCT  60

Qy    1189 GTGTTCCTGGGAGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCACGTTCCT 1248
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTCTGCTGGGGGAAGGATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCACGTTCCT 120

Qy    1249 CGTTGCACCAAGCCCAGCAGAGGCCCAGACTCCAGAGATGAGCTCTTCAGCCAGCTGGTGC 1308
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGTTAGCAGGTCCAGGGATGGTGGGCAGTTCCAGATGAGCTCTTCAGCCAGCTGGTGC 180

Qy    1309 CTATTGTGTGGGCAAACCCAGATGAGCAACAGAATCAGCCTGGCTGGGCCCTAATGGTGAT 1368
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTAGTTGTGGGCAAACCCAGATGAGCAACAGAATCAGCCTGGCTGGGCCCTAATGGTGAT 240

Qy    1369 CCTGTTCAGCTCCTTGCTGCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATC 1428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCTGTTCAGCTCCTTGCTGCCACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATC 300

Qy    1429 TGACTAGCCTCCCAGGTGATGGTGCCAGCTACAAGCTGTTAGGTGCCCTGGA 1478
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGACTAGCCTCCCAGGTGATGGTGCCAGCTACAAGCTGTTAGGTGCCCTGGA 350

Qy    1479 GACAGACAGCTGGCCTCCATGGGTTTGAGGTCCCACCCACCCACACAACTTGAGTGGAA 1538
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 GACAGACAGCTGGCCTCCATGGGTTTGAGGTCCCACCCACCCACACAACTTGAGTGGAA 410

Qy    1539 GTTTGGTTTAGTCCGGGCCCGCATGGGCTGGATGTGTTCACATTCAACGAGGAAAGCTA 1598
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 GTTTGGTTTAGTCCGGGCCCGCATGGGCTGGATGTGTTCACATTCAACGAGGAAAGCTA 470

Qy    1599 CTTTTGGAAGTGGAATCCTGGACTACGGGAGAGCAGTTTGCAGGGTGGATCCTACAGAG 1658
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     471 CTTCTGGAAGTGGAATCCTGGACTACGGGAGAGCAGTTTGCAGGGTGGATCCTACAGAG 530

Qy    1659 CAGAGGCCTGGAGGCGCCCCCTCGTGGCTGGTCTGTGTCACTGCATTCTGGGGATGCTTG 1718
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 CAGAGGCCTGGAGGCGCCCCCTCGTGGCTGGTCTGTGTCACTGCATTCTGGGGATGCTTG 590

Qy    1719 GTGTACTTGCCTGGCTGTGACTTTCTGTTTGAGCCTAATAGGCCAGACTGAGGACTTGGG 1778
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     591 GCCTGACTTGCCTGGCTGTGACTTTCTGTTTGAGCCTAATAGGCCAGACTGAGGACTTGGG 650

Qy    1779 AGACTTAGCTGGCCCCCAACAACTAGCCCATCACTCCTCTTGGTTTAGCTGAGAGCATCCC 1838
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     651 AGCCTCAGCTGTCCCCCAACAACTAGCCCATCACTCCTCTTGGTTTAGCTGAGAGCATCCC 710

Qy    1839 TTTAGTTTTGGTGTCTAGGTTCTTTCCTGCCCCAGGACTCCCTCCAGGTCCAGCCCC 1898
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     711 TTTAGTTTTGGTGTCTAGGTTCTTTCCTGCCCCAGGACTCCCTCCAGGTCCAGCCCC 770

Qy    1899 AGTATTGGCCAGCAGCCCCCCCT-CCGGGCAGCCAGTAAGCCTCAGAAGCCTGGATGGTT 1957
           ||| |||||||||||||||||||| ||||||||||||||||||| |||||||||||||
Db     771 AGTCTTGGCCAGCAGCCCCCCCTCCCGGGCAGCCAGTAAGCCTCAGAAGCCTGGATGGTT 830

Qy    1958 GGGCTTAGCTGTGGGAACAAGTAAGTTTCTGAAAGCTGGGATGGGT 2003
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     831 GGGCTTAGCTGTGGGAACAAGTAAGTTTCTGAAAGCTGGGATGGGT 876

Qy    1908 TCGTGGACCACCTCTTTGAACTAGCGTTGCTTCC 1941
           || |||||| |||||| | |||||||||| ||||
Db     841 TCTTTGACTACTTTTTAATAGCGTTGGTTCCG 874


RESULT 2
BQ956737
LOCUS       BQ956737                  924 bp    mRNA    linear   EST 17-APR-
DEFINITION  AGENCOURT_6399024 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6399024 5', mRNA sequence.
ACCESSION   BQ956737
VERSION     BQ956737.1  GI:22372215
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   NIH MGC https://mgc.ncbi.nlm.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cooke Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13899  row: h  column: 01
            High quality sequence stop: 651.
FEATURES             Location/Qualifiers
     source          1..923
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6399024"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      235 a    243 c    209 g    183 t      2 others
ORIGIN

  Query Match             14.2%;  Score 898.6;  DB 14;  Length 924;
  Best Local Similarity   96.0%;  Pred. No. 1.4e-139;
  Matches  889;  Conservative    0;  Mismatches   41;  Indels    9;  Gaps    4;

Qy    4780 TAAGGGGGGGGGGAAGGAGGAGGTGAGGTGCTCATGAGGGGGGAAGGAGGAAGCCCCAGGT 4839
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TAAGGCCCCGGGGCAAGGACGAGGTGAGGTGCTCATGAGGGGGGAAGGAGGAAGCCCCAGGT  60

Qy    4840 TGACCTTAGGGACAGATGTACTGCCAGGTCATCAAGGAAGTGACAGAGACCCCTGCCTGCT 4899
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGACCTTAGGGACAGATGTACTGCCAGGTCATCAAGGAAGTGACAGAGACCCCTGCCTGCT 120

Qy    4900 AAAGCACTGTGCTGTGGGTGGACCCTGTGAGCCTGTGAGCCTGTGCAGGCTGTTTGCACCAT 4959
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AAAGCACTGTGCTGTGGGTGGACCCTGTGAGCCTGTGAGCCTGTGCAGGCTGTTTGCACCAT 180

Qy    4960 GACCACGCTGATGCCCTATGTGACCAAGTTCCTGCAGGATCCAGCCCTAGTGAAGTTT 5019
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GACCACGCTGATGCCCTATGTGACCAAGTTCCTGCAGGATCCAGCCCTAGTGAAGTTT 240

Qy    5020 GGCCAGGAGGAGCCAGGAGAACCTCCAGCCAGCAGTAAATATCCCCAGCCAGCTGTT 5079
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCCAGGAGGAGCCAGGAGAACCTCCAGCCAGCAGTAAATATCCCCAGCCAGCTGTT 300

Qy    5080 GGCGTTACCTCTGGAAAAGAAGTCTTTCTGAAAGCAGCAGCAGTTCTTCCTCTAAT 5139
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGCGTTACCTCTGGAAAAGAAGTCTTTCTGAAAGCAGCAGCAGTTCTTCCTCTAAT 360
```

```
Db     301 GCCGTTACCTGGTGAAATGAATGCTTTTCTGAAAGGGCAAGCAGTTCGTTTGCTTCTAAT 360

Qy    5140 TCACCTGCCTGGGGGTGTGGACTACAAGACGAATTCACAGACATTCACAGTGGCAGGGGA 5199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCACCTGCCTGGGGGTGTGGACTACAAGACGAATTCACAGACATTCACAGTGGCAGGGGA 420

Qy    5200 AGTGCTAGAGGAGCTGTGTGGACAGATGGGCATCACAGACTTGGAAGAAGTGCAGGAATT 5259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGTGCTAGAGGAGCTGTGTGGACAGATGGGCATCACAGACTTGGAAGAAGTGCAGGAATT 480

Qy    5260 TGCCCTCTTTCTTCATCAAAGGAGAAGGTGAGCTGGTTGGGCGGCTGTCACCCCATGAGTA 5319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGCCCTCTTTCTTCATCAAAGGAGAAGGTGAGCTGGTTGGGCGGCTGTCACCCCATGAGTA 540

Qy    5320 CATCAACAATGTGGTGACGGACCAGGACATGAGCCTTCACAGCCGACGGCTTGGTTGGGA 5379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CATCAACAATGTGGTGACGGACCAGGACATGAGCCTTCACAGCCGACGGCTTGGTTGGGA 600

Qy    5380 GAGTCTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGACGAGATGTACTG 5439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAGTCTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGACGAGATGTACTG 660

Qy    5440 GGACTACCTGCAAGGCAAGTTGATAGTCAAGCCCAGGCAGAGGCTCTACTTGCCCAGCT 5499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGACTACCTGCAAGGCAAGTTGATAGTCAAGCCCAGGCAGAGGCTCTACTTGCCCAGCT 720

Qy    5500 TGCTGCCTTCCAACAGTTCCAGAAAACGGGAACTTCTAGTCCTCCATCAGAGCAAGAGCT 5559
           ||||| ||| | ||||||||||||||| ||||||||||||| ||||||||||||||||||||
Db     721 TGCTG CTTTCACACTTCACAAAACGGGAACTTCTAGTCCTTATCAGAGCAAGAGCT 778

Qy    5560 GCTGTCTTAIATCCCCAAGCACTGCAATGGCAGGTGAACACAGCCAACATAAAGAGCTT 5619
           |||||||||||||||||| |||||||||||   ||| |||||| ||     ||||||||
Db     779 GCTGTCTTATATTCCCAAG-CACTGCAATGGNCAGTG ACACAGNCACATAAAAGAGCTG 837

Qy    5620 GGTGACCCAGGAGCTGAGCCAGATGTAAGGGTAAGCAAGCAGAGAGCACAGA --TTGG 5676
           | |||||| | |||||||||||| | ||||||||||| |  ||          ||
Db     838 GNTGANCCCNGNAGCTNANGCANATGGCAANGCAGTAAGCANAAGAGCAACAGATGGNG 897

Qy    5677 CTTTATAGAGAGCACAGCGCAGCTGC 5703
           |||||||||||||||||||||||||||
Db     898 CTTTATAGAGAGCACAGCGCAGCTGC 923


RESULT 3
BG968181
LOCUS       BG968181                   881 bp    mRNA    linear   EST 12-JUN 2001
DEFINITION  CGAP Mus musculus cDNA clone IMAGE:4990215 5',
            mRNA sequence.
ACCESSION   BG968181
VERSION     BG968181.1  GI:14356818
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11065  row: k  column: 16
            High quality sequence stop: 838.
FEATURES             Location/Qualifiers
     source          1..881
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4990215"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      225 a    233 c    242 g    181 t
ORIGIN

  Query Match          12.6%;  Score 700.2;  DB 13;  Length 881;
  Best Local Similarity   97.6%;  Pred. No. 3.7e-123;
  Matches  866;  Conservative    0;  Mismatches   13;  Indels    8;  Gaps    6;

Qy    4744 CTTCAAGGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCGGGGCAAGGACGAGCT 4803
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTTCAAGGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCGGGGCAAGGACGAGCT 60

Qy    4804 GAGTCTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGACGAGATGTACTG 4863
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAGTCTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGACGAGATGTACTG 120

Qy    4864 CCAGGTCATCAAGCAAGTCACAGGACACCCCAGCAAAGCACTGTCTTCTGGCTGGAG 4923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCAGGTCATCAAGCAAGTCACAGGACACCCCAGCAAAGCACTGTCTTCTGGCTGGAG 180

Qy    4924 CGTCCTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACCAGCGTGATGCCCTATGTGAC 4983
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGTCCTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACCAGCGTGATGCCCTATGTGAC 240

Qy    4984 CAAGTTCCTGCAGGATTCCAGCCCCAGTCAAGAGTTCTCCAGGAGGAGCCAGCAGAACCT 5043
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAAGTTCCTGCAGGATTCCAGCCCCAGTCAAGAGTTCTCCAGGAGGAGCCAGCAGAACCT 300

Qy    5044 CCAGCGCACAGTTAAATATGGGGGACGCCAGCAGCTGCCGTTACCTGGTGAAATGAATGC 5103
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCAGCGCACAGTTAAATATGGGGGACGCCAGCAGCTGCCGTTACCTGGTGAAATGAATGC 360

Qy    5104 TTTTCTGAAAGGGCAAGCAGTTCGTTTGCTTCTAATTCACCTGCCTGGGGGTGTGGACTA 5163
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTTTCTGAAAGGGCAAGCAGTTCGTTTGCTTCTAATTCACCTGCCTGGGGGTGTGGACTA 420

Qy    5164 CAGGACGAATTCACAGACATTCACAGTGGCAGGGGAAGTGCTAGAGGAGCTGTGTGGACA 5223
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAGGACGAATTCACAGACATTCACAGTGGCAGGGGAAGTGCTAGAGGAGCTGTGTGGACA 480

Qy    5224 GATGGGCATCACAGACTTGGAAGAAGTGCAGGAATTTGCCCTCTTTCTTCATCAAAGGAGA 5283
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GATGGGCATCACAGACTTGGAAGAAGTGCAGGAATTTGCCCTCTTTCTTCATCAAAGGAGA 540

Qy    5284 AGGTGAGCTGGTTGGGCGGCTGTCACCCCATGAGTACATCAACAATGTGGTGACGGACCA 5343
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AGGTGAGCTGGTTGGGCGGCTGTCACCCCATGAGTACATCAACAATGTGGTGACGGACCA 600

Qy    5344 GGACATGAGCCTTCACAGCCGACGGCTTGGTTGGGA--GAGTCTGCTATGAGCTGCTGAA 5401
           |||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db     601 GGACATGAGCCTTCACAGCCGACGGCTTGGTTGGGAAGAAGTCCACTGCA-TTTGATCAC 659

Qy    5402 TCCAGCTACAGCGGAAAGCCACTATGGCCAGGTGCTTCGGGACTACCTGCAAGGGAAGCTG 5461
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 TCCAGCTACAGCGGAAAGCCACTATGGCCAGGTGCTTCGGGACTACCTGCAAGGGAAGCTG 719

Qy    5462 ATAGTCAGCAGCCCAGGCAGAGGCCTCTACTTGCCCAGCTTCCTCCAACAGTTCCAAC 5521
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 ATAGTCAGCA-CCAGGCAGAGGCTCTACTTGCCCAGCTTCCTCC--TTCAAACTTCGAA 776

Qy    5522 AAAACGGGAACTTCTAGTCCTCCATCAGAGCAAGAGCTGCTGTCTTATATCCCCAAGCCA 5581
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     777 AAAACGGGAACTTCTAGTCCTCCATCAGAGCAAGAGCTG-TGTCTTATATT-CCAAGCCA 834

Qy    5582 CTGCAATGGCAGGTGAACACAGCCAACATAAAGAGCTTGGTGACCCA 5628
```

```
Db    835 CTCAAGCCAGGGTGAACATAGCACACATAAAGAGCATGGTGAAACCA 881


RESULT 4
BQ960468
LOCUS       BQ960468                  925 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8950845 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6474896 5', mRNA sequence.
ACCESSION   BQ960468
VERSION     BQ960468.1  GI:22375946
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14012  row: m  column: 09
            High quality sequence stop: 671.
FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6474896"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally  Primer: Oligo dT.
                     Average insert size 1.7 Kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      174 a    287 c    290 g    174 t
ORIGIN

  Query Match          13.3%;  Score 274.2;  DB 14;  Length 925;
  Best Local Similarity   95.0%;  Pred. No. 1.8e-100;
  Matches  361;  Conservative    0;  Mismatches   28;  Indels    9;  Gaps    6;

Qy      781 ATCCTTTGCGGACAGCGGGGAGCCCTGGGGGAAAGTGTCAAATATGGACCTGGGTCGC 840
            |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATCCTTGGCCGCACAGCGGGGAGCCCTGGGGGAAAGTGTCAAATATGGACCTGGGTCGC 60

Qy      841 TTAGAGAAGGCGCGCAGCTAGCTACTCTGCTGGAGAGGGCGGAAGGCCACCAGGCCTTG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 TTAGAGATCCCCCGCCAGCTAGCTACTCTGCTGGAGAGGGCGGAAGGCCACCAGGCCTTG 120

Qy      901 CTAAGAGGAGCATCACAGATCCCTGCCAGTGAGGTCCCGGCCGGCCAGCCTGACT 960
            ||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CTAAGAGGAGCATCACAGATCCCTGCCAGTGAGGTCCCGGCCGGCCAGCCTGACT 180

Qy      961 CTCCT--AGAGATGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCAGAAG 1018
            |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CTCCTTAGAGATGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCAGAAG 240

Qy     1019 CTATTGTGGCTGGACCAGGGCAGCCACTGGACGAGCCCCTGACGGCGTTAGATGGCGAG 1078
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CTATTGTGGCTGGACCAGGGCAGCCACTGGACGAGCCCCTGACGGCGTTAGATGGCGAG 300

Qy     1079 AAGCTTCAGCAGGCTCTGGAGATCAACAGGGTGATGCTGCGGCTCCTGGGGGAAGGATCT 1138
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AAGCTTCAGCAGGCTCTGGAGATCAACAGGGTGATGCTGCGGCTCCTGGGGGAAGGATCT 360

Qy     1139 GCAGAGGTGTGTGGCAGCACAAGCTGTTAGGTGCCTGAGCACACACTGGCTCCCATC 1198
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GCAGGCCTGTGCCAGCACAAGCTGTTAGGTGCCTGAGCACACACTGGCTCCCATC 420

Qy     1199 CTGCAGTCCTGGCAAGAGCAGACATGGGCACCCTCCTCCCAGTCCTCAGTCCATGAGT 1258
            |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 CTGCAGTCCTGGCAAGAGCAGACATGGGCACCCTCCTCCCAGTCCTCAGTCCATGAGT 480

Qy     1259 CCGGGACTCCGACATCAGCCTCTTCAGCCAGCTGGTGGCCACCATCTGTGTGCCTAAGTAT 1318
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CCGGGACTCCGACATCAGCCTCTTCAGCCAGCTGGTGGCCACCATCTGTGTGCCTAAGTAT 540

Qy     1319 GAGCAACAGAATCAGCGTGGCTGGGCCTAATGGTGATCCTGCTGAGTCTTTCTCTCTC 1378
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GAGCAACAGAATCAGCGTGGCTGGGCCTAATGGTGATCCTGCTGAGTCTTTCTCTCTC 600

Qy     1379 ACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATTTGACTGCCCAGTGGTATTCCCC 1438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 ACACCTGCCCTGGAGAAGCCACTGCTCAAATTTGTATTTGACTGCCCAGTGGTATTCCCC 660

Qy     1439 GCAGCCCTGTGCCAGCACAAGCTGTTAGGTGCCTGAGCACACACTGGCTCCCATC 1498
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GCAGCCCTGTGCCAGCACAAGCTGTTAGGTGCCTGAGCACACACTGGCTCCCATC 720

Qy     1499 GCCTCG-AGGTCCCACCCACCCACACAAGCTTGGAGTCCCAGGTCCCCCTCCCCCTCCCC 1557
            ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 CGCATGGGGGCTGGATGTGTTCACATTCAACGAGGAAAGTCACTCCCCCAAGTCCAACTC 780

Qy     1558 CTGGACCACGGGAGAGCCAG-TTTGCAGGGTGGATCCTACAGAGCAGA-GGCCTGGAGGC 1615
            ||:|||||||||||||||| ||    ||||||||||||||||||||| || | ||||||||
Db      781 CTGGACCACGGGAGAGCCAGGTTTGCAGGGTGGATCCTACAGAGCAGAGGGCCTGGAGGC 840

Qy     1616 GCCCCC--TCGTGGCTGGTCTGTGTGCACTGCAGTCCTGGGGAGTGCTTGGGGTGGACTTG 1673
            ||||||  |  |||||||||| |||||||  |   |  |  | | |  |  |  |||||
Db      841 GCCCCCTTGCTGGGCTGGTCTGTGTGCACTGCAGTCCTGGGGAGTGCTTGGGGTGGACTTG 900


RESULT 5
BG962322
LOCUS       BG962322                  877 bp    mRNA    linear   EST 1-JUN-2002
DEFINITION  AGENCOURT_8950845 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:6474896 5',
            mRNA sequence.
ACCESSION   BG962322
VERSION     BG962322.1  GI:14340959
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Jeffrey F. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10984  row: m  column: 05
            High quality sequence stop: 773.
FEATURES             Location/Qualifiers
     source          1..877
                     /organism="Mus musculus"
                     /strain "FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4982188"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
```

```
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally  Primer: Oligo dT
                     Average insert size 1.6 kb  Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library "
BASE COUNT      221 a    220 c    244 g    186 t
ORIGIN

  Query Match             11.6%;  Score 721.8;  DB 13;  Length 877;
  Best Local Similarity   97.5%;  Pred. No. 1.2e-111;
  Matches  308;  Conservative    3;  Mismatches   11;  Indels    3;  Gaps    3;

Qy   5054 AGTTAAATATGGGGGACGGCAGCAGCTGCCGTTACCTGGTGAAATGAATGCTTTTCTGAA 5113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGTTAAATATGGGGGACGGCAGCAGCTGCCGTTACCTGGTGAAATGAATGCTTTTCTGAA 60

Qy   5114 AGGGCAAGCAGTGGTTTGTTCTAATTACACTTGCCTGGGGGTGTGGACTACAGGACGAA 5173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGGCAAGCAGTGGTTTGTTCTAATTACACTTGCCTGGGGGTGTGGACTACAGGACGAA 120

Qy   5174 TTCACAGACATTCACAGTGGCAGGGAAGTGCTACAGGAGCTGTGTGGACAGATGGGCAT 5233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTCACAGACATTCACAGTGGCAGGGAAGTGCTACAGGAGCTGTGTGGACAGATGGGCAT 180

Qy   5234 CACAGACTTTGGAAGAAGTGTACGAATTTGCCCTCTTTTCATCAAAGGAGAAGGTGAGCT 5293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACAGACTTGGAAGAAGTGTACGAATTTGCCCTCTTTTCATCAAAGGAGAAGGTGAGCT 240

Qy   5294 GGTTCGGCCGCTGTCACCCCATGATACATCAACAATGTGGTGACGGACCAGGACATGAG 5353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGTTCGGCCGCTGTCACCCCATGATACATCAACAATGTGGTGACGGACCAGGACATGAG 300

Qy   5353 CCTTCACAGCCGACGGCTTGGTTGGGAGACTCCACTGCATTTTGATCACTCCACCTACAC 5412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTTCACAGCCGACGGCTTGGTTGGGAGACTCCACTGCATTTTGATCACTCCACCTACAC 360

Qy   5413 GGAAACCCACTATGGCCAGCTGTCTTCGGACTACTTCCAAGGGAAGCTGATAGTCAGCAC 5472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAAACCCACTATGGCCAGCTGTCTTCGGACTACTTCCAAGGGAAGCTGATAGTCAGCAC 420

Qy   5473 CCAGGCAGAGGCTCTACTTGCCCAGTTGCTGCCTTCCAACACTTCGACAAAACCGGAAAC 5532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCAGGCAGAGGCTCTACTTGCCCAGTTGCTGCCTTCCAACACTTCGACAAAACCGGAAAC 480

Qy   5533 TTCTAGTCCTCCATCAGAGTAAGAGCTGCTGTCTTATATTCCCAAGCCACTGCAATTGGC 5592
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    481 TTCTAGTCCTCCATCAGAGTAAGAGCTGCTGTCTTATATTCCCAAGCCACTGCAATTGGC 540

Qy   5593 AGGTGAACACAGCCAACATAAAGAGCTTTGGTGACCCAGGAGCTGAGGCAGATGCAAGGG 5652
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    541 AGGTGAACACAGCCAACATAAAGAGCTTTGGTGACCCAAGGAGCTGAGGCAGATGCAAGGG 600

Qy   5653 TACAGCAAGCAGAGAGTACAAATTTCTTTATAGAGAGCACAGCGGTAGTGCCCCTCTTT 5712
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TACAGCAAGCAGAGAGTACAGATTGG TTTATAGAGAGCACAGCGGTAGTGCCCCTCTTT 659

Qy   5713 GGCTACAC TCTCTACTTAGTG-CTGAGAGTG-AGTAAGTTGGCCCTCCCT GGAGCAGG 5766
          |||||||| |||||||||||||| |||||||| |||||||||| ||||||| ||||||||
Db    660 GGCTACACTTGTGTACGTAGTGCCTGAGAGTGAAGTAAGTTGGACCTCCCTGGGACCAGG 719

Qy   5767 CCTCCTGGGGGTTGAAGTT-TTCAGCAGCTGGTTCTCATGGA-CCCCAGCTCTCAGGAAGTC 5824
          |||||||||||||||||| |||||||||||||||||||||| ||||| |||||||||||||
Db    720 CCTCCTGGGGGTTGAAGTTGTTCAGCACTTGGTTCTCATGGACCCCTAAGTTCTCAGGGACTC 779

Qy   5825 TGCCTGCTCTGTCATGCTAAAAAGACCTGAAGTAGTTCCACCTGCTGAGCCCACTGCAGGAG 5884
          |||||||||||||||||| |||||||| ||| ||||| |||||||||| |||| |||||||
Db    780 TGCCTGCTCTGTCATGCTTAAAATA TTTGAAACAAGTCCACTTGCTGAGCCAATGCATGAG 839

Qy   5885 GACGGGCCC 5893
          |||||||||
Db    840 GACGGGCCC 848

RESULT 6
```

```
BQ941605
LOCUS       BQ941605                 947 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8979776 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6475749 5', mRNA sequence.
ACCESSION   BQ941605
VERSION     BQ941605.1  GI:22357083
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14014  row: p  column: 22
            High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6475749"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb  Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      175 a    319 c    290 g    161 t      2 others
ORIGIN

  Query Match             10.8%;  Score 680;  DB 14;  Length 947;
  Best Local Similarity   95.2%;  Pred. No. 5.7e-105;
  Matches  760;  Conservative    0;  Mismatches   30;  Indels    9;  Gaps    6;

Qy    787 GTGTGCCCCTGGCCGCACAGCGGGGAGCCTGGGGAAAGTCCCAAATATGGACCTGGG 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     32 GTGTGCCCCTGGCCGCACAGCGGGGAGCCTGGGGAAAGTCCCAAATATGGACCTGGG 91

Qy    847 TGGCTTAGAGATCCCCGGCCAGCTGGCTACTCTGCTGGAGAGAGGGAAGCCAACCAGGC 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 TGGCTTAGAGATCCCCGGCCAGCTGGCTACTCTGCTGGAGAGAGGGAAGCCAACCAGGC 151

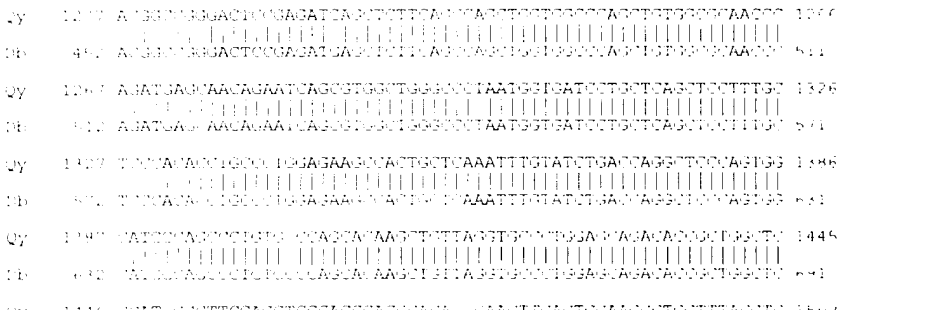
Qy    907 CTTGTTGACGGGGAGCATCACAGAGTCCCTGCCACCTGAGGTCCCCCCCCCCCAGCCT 966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 CTTGTTGACGGGGAGCATCACAGAGTCCCTGCCACCTGAGGTCCCCCCCCCCCAGCCT 211

Qy    967 GACTCTCCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCA 1026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 GACTCTCCCCTCCAGACATTGACCAGTTTCCCTTCTCCAGTTTTGTATCCACCAGCTTTCA 271

Qy   1027 GAAGCCATTTCTGCCTCGACCAGGGCAGCCACTGGACGAGCCCCTGACCGGGTTAGATGG 1086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 GAAGCCATTTCTGCCTCGACCAGGGCAGCCACTGGACGAGCCCCTGACCGGGTTAGATGG 331

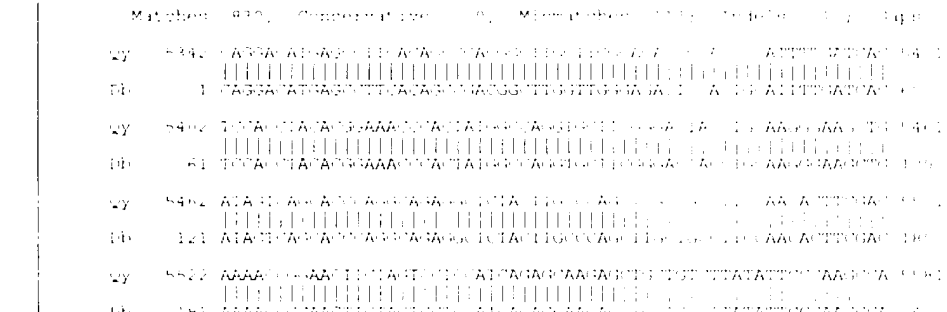
Qy   1087 CGAGAACCCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCTGTGGCTCCTGGGGGAAGG 1146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 CGAGAACCCTCAGCAGGCTCTGGAGATCAACAGGGTGATGCTGTGGCTCCTGGGGGAAGG 391

Qy   1147 ATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCAGTTCCTCGTGCAGCAGGCCCAGCG 1206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    392 ATCTCTGCAGTCCTGGCAAGAGCAGACCATGGGCAGTTCCTCGTGCAGCAGGCCCAGCG 451
```



RESULT 7
BF584426
LOCUS BF584426 1006 bp mRNA linear EST 12-DEC-2000
DEFINITION [illegible] mRNA sequence.
ACCESSION BF584426
VERSION BF584426.1 GI:11656138
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1006)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM9773 row: c column: 20
High quality sequence stop: 684.
FEATURES Location/Qualifiers
source 1..1006
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4208587"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Vector: pCMV-SPORT6; Site_1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 295 a 288 c 249 g 174 t
ORIGIN

Query Match 10.6%; Score 670.2; DB 12; Length 1006;
Best Local Similarity 86.9%; Pred. No. 5.4e-103;



RESULT 8
BF584026
LOCUS BF584026 [illegible] bp mRNA [illegible] EST 12-DEC-2000
DEFINITION [illegible] mRNA sequence.
ACCESSION BF584026

```
VERSION     BF584026.1  GI:11657744
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph D.
            Email: cgapbs-rmail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM3703  row: e  column: 17
            High quality sequence stop: 690.
FEATURES             Location/Qualifiers
     source          1..932
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4216312"
                     /clone_lib "NCI_CGAP_Ov24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon, Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      220 a    244 c    294 g    174 t
ORIGIN

  Query Match             10.5%;  Score 658;  DB 12;  Length 932;
  Best Local Similarity   98.0%;  Pred. No. 6.2e-161;
  Matches  649;  Conservative    7;  Mismatches   10;  Indels    4;  Gaps    3;

Qy    3241 GGACTGCGGCTGCTGAAGGTGACCCAAAGCCCGAGTTCCAGCCTGGACCAGCTGAAGACA 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGACTGCGGCTGCTGAAGGTGACCCAAAGCCCGAGTTCCAGCCTGGACCAGCTGAAGACA 60

Qy    3301 CTCTGTTCCTACAGTCATGTTGAAGTCCTGACCGTGCAGTGCAGGGGCAGATCCACCCTG 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTCTGTTCCTACAGTCATGTTGAAGTCCTGACCGTGCAGTGCAGGGGCAGATCCACCCTG 120

Qy    3361 GAGTGGTGTTTGAAGAATGAATAGCTGATATTGCACACAGCCTGGGCGAGGGCCATCAAG 3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAGCTGTCCTTGAAGAATGAATAGCTGATATTGCACACAGCCTGGGCGAGGGCCATCAAG 180

Qy    3421 GCCATGGTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTCATCGCCCTG 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCATGGTGGATCTATTTCTGAGTGAACTCAGGAAGGACTCCGGCTATGTCATCGCCCTG 240

Qy    3481 CGCAGCTACAICAGCGATGACAATAGCCTCCTCAGTTTCCACCGTGGGGACCTCATTAGG 3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGCAGCTACATCAGCGATGACAATAGCCTCCTCAGTTTCCACCGTGGGGACCTCATTAGG 300

Qy    3541 TTACTGCCAGTGACCGTTCTGGAAGCTAGGCTGGCAGTTCGGTTCTGCCGGGGGCCGCTCC 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTACTGCCAGTGACCGTTCTGGAAGCCAGGCTGGCAGTTCGGTTCTGCCGGGGGCCGCTCC 360

Qy    3601 GGACTCTTTCCCGATGACCTGGTGTACCTAGCTGTTGCCCCGGACCTCTCCTTTTCCCTG 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGACTCTTTCCCGATGACCTGGTGTACCTAGCTGTTGCCCCGGACCTCTCCTTTTCCCTG 420

Qy    3661 GGAAAGAGAAACAGCTTGGCAACTGCAAGAGTAAGCTGGGGCCAGCTCAGGAGGTGAGGAAG 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAAAGAGAAACAGCTTGGCAACTGCAAGAGTAAGCTGGGGCCAGCTCAGGAGGTGAGGAAG 480

Qy    3721 ACAGAAGAGGTGAAGTGATACAGGCCTAACTTGGAGACTGAGAAGGAAAGAGCAGGGTTG 3780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACAGAAGAGGTGAAGTGATACAGGCCTAACTTGGAGACTGAGAAGGAAAGAGCAGGG-TG 538

Qy    3781 CTTCGGGTGTTGTCCACTTCCTGTCCTGGTGGCCAGGGCTCAATGTGTTCCTGTCCTTTA 3840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 CTTCGGGTGTTGTCCACTTCCTGTCCTGGTGGCCAGGGCTCAATGTGTTCCTGTCCTTTA 598

Qy    3841 CCATCTCCTGACTTTTTGCCATTTGTGAGACTGTAAGTCACACCCTCTAACTCTGGTACT 3900
           ||||||||||||||||  |  ||||||||||||||||||||||||||||||||||||||||
Db     599 CCATCTCCTGACTTTT--GCCATTGTGAGACTGTAAGTCACACCCTCTAACTCTGGTACT 656

Qy    3901 TAGTTCAGTGTCTCCATAGAGGATGCTTAATAAATAACCTTGGTTTTCCTGG 4352
           ||||||    |||||||||||||||||||||||||||||| | ||||||||||
Db     657 TAGTTTCAGGTCTCCATAGAGGATGCTTAATAAATAACCTTGGTTTTCCTGG 708


RESULT 9
BB072116
LOCUS       BB072116                 677 bp    mRNA    linear   EST 18-OCT-2001
DEFINITION  BB072116 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030008I18 3', mRNA sequence.
ACCESSION   BB072116
VERSION     BB072116.2  GI:16260140
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 19, 2000 this sequence version replaced gi:8582114.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             Wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Arizawa
            ,F., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
```

```
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN,
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030008I18"
                     /clone_lib="RIKEN full length enriched, adult male colon"
                     /sex="Male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI, Site_2: BamHI, cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap trapper. cDNA went through one round of normalization
                     to Rot -/- 20.0 and subtraction to R-/- 195.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATTCCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       148 a    176 c    186 g    165 t      2 others
ORIGIN

  Query Match             10.4%;  Score 657.6;  DB 10;  Length 677;
  Best Local Similarity   99.0%;  Pred. No. 7.6e-101;
  Matches  ###;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy    4671 AGTTTCCACTTGGACCAGATGAAGACACTCTGTTCCTACAGTTATG TGAAGTCCTGACC 4844
           ||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db       1 AGTTTCCACTTGGACCAGATGAAGACACTCTGTTCCTACAGTTATG TGAAGTCCTGACC 59

Qy    4845 TCTCAGTGCAGGGCCAGATCCACCCTGGAGCTGTCCTTGAAGAATGAGCAGCTGATACTG 3793
           ||| | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 TCTGAGTGCAGGGCAGATCCACCCTGGAGCTGTCCTTGAAGAATGAGCAGCTGATACTG 119

Qy    3792 CACACAACCTCCCCCAGGCCATCAAGGCCATGGTGGATCTATTTCTGAGTGAACTCAGG 3853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 CACACAGCCTGGGCGAGGGCCATCAAGGCCATGGTGGATCTATTTCTGAGTGAACTCAGG 179

Qy    3852 AAGTAATTCCCCCTAATCCATCGCCCTGGGTAGTACATCACCGATGATAATAGCCTCTCC 3673
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 AAGTAATTCCGGTTATCCTCATCGCCCTGGGTAGNTACATCACCGATGATAATAGCCTCTCC 239

Qy    3672 AGTTTCCACCGTGGGGACCTCATTAGGTTACTGCCAGTGACCGCTCTGGAACCAGGCTGG 3493
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 AGTTTCCACCGTGGGGACCTCATTAGGTTACTGCCAGTGACCGCTCTGGAACCAGGCTGG 299

Qy    3492 CAGTTCGGTTCTGCCGGGGGCCGCTCCGGACACTCTTCCCGATGACAGCCCTGAGCCAG 3313
           |||| | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 CAGTTTGGTTCTGCCGGGGGCCGCTCCGGACACTCTTCCCGATGACAGCCCTGAGCCAG 359

Qy    3312 CTTCCCCCCACCTCTCCTTTCCCTGGAAAGAGAAACAGTTGGTAAAGTAAAAGTAAAA 3133
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 CTTCCCCCCACCTCTCCTTTCCCTGGAAAGAGAAACAGTTGGTAAAGTAAAAGTAAAA 419

Qy    3132 TTGGGGCCAGCTCAGGAGGTGAGGAAGACAGAAGAGGTGAAGTGATACAGGCCTAACTTG 2953
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 TTGGTGCCAGCTCAGGAGGTGAGGAAGACAGAAGAGGTGAAGTGATACAGGCCTAACTTG 479

Qy    2952 AGAATTGAGAAGGAAAGAGCAGGGTTGCTTGGGGTGTTGTCCACTTCCTGTCCTGATGT 2773
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 AGAATTGAGAAGGAAAGAGCAGGGTTGCTTGGGGTGTTGTCCACTTCCTGTCCTGATGT 539

Qy    4214 CAGGGCTCAATGTGTTCTGTCTTTACATCTCTAATTTTTTTATTTGTGAAATTTGATG 4033
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 CAGGGCTCAATGTGTTCCTGTCCTTTACATCTCTAATTTTTTTATTTGTGAAATTTGATG 599

Qy    4032 TAAGTCACACCCTCTAACTCTGGTACTTAGTTCAGTGTCTTTATAGAGATGCTTAATAA 3853
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 TAAGTCACACCCTCTAACTCTGGTACTTAGTTCAGTGTCTTTATAGAGATGCTTAATAA 659

Qy    3852 ATAACCTTGGTTTTCCTG 3835
           ||||||||||||||||||
Db     660 ATAACCTTGGTTTTCCTG 677


RESULT 10
BG962743
LOCUS       BG962743                 669 bp    mRNA    linear   EST 15-JUN-2
DEFINITION  [illegible] Mus musculus [illegible] cDNA clone IMAGE:[illegible],
            mRNA sequence.
ACCESSION   BG962743
VERSION     BG962743.1  GI:14350389
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaphs@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium, LLNL
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10991  row: d  column: 17
            High quality sequence stop: 669.
FEATURES             Location/Qualifiers
     source          1..669
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4984672"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: Not I;
                     Site_2: Sal I, Cloned unidirectionally.  Primer: 3' CAP
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP Library."
BASE COUNT       167 a    194 c    172 g    136 t
ORIGIN

  Query Match             10.4%;  Score 657;  DB 13;  Length 669;
  Best Local Similarity   99.9%;  Pred. No. 9.9e-101;
  Matches  668;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy    4671 AAAGAAGTTAGTAAGAATTTTTATTTTATTTTTTTTATTTTTTTTTTTTTTTTTTT 4850
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAAGAAGTTAGTAAGAATTTTTATTTTATTTTTTTTATTTTTTTTTTTTTTTTTTT 60

Qy    4851 GAATTTTATAAATAATATTTATGTATAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5030
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAATTTTATAAATAATATTTATGTATAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 120

Qy    5031 TTTTTATTTTATTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5210
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGTTATTTTATTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 180
```

```
Qy    4679 ACTAAGGACCCCATCCAGGAATCCCTTACCAGCTTCTGCAATGGGGACACAAACAGTAAA 4738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACTAAGGACCCCATCCAGGAATCCCTTACCAGCTTCTGCAATGGGGACACAAACAGTAAA 240

Qy    4739 GCTGTGGCTGGCTTCAAGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCGGGGC 4798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCTGTGGCTGGCTTCAAGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCGGGGC 300

Qy    4799 AAGGACGAGCTGAGTGTTGCTGTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGAC 4858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAGGACGAGCTGAGTGTTGCTGTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGAC 360

Qy    4859 GAGATGTACTGCCAGGTCATCAAGCAAGTCACAGGACACCCCCAGCCAAAGCACTGTGCT 4918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAGATGTACTGCCAGGTCATCAAGCAAGTCACAGGACACCCCCAGCCAAAGCACTGTGCT 420

Qy    4919 CTGGGCTGGAGCGTCCTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACCACGCTGATG 4978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CTGGGCTGGAGCGTCCTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACCACGCTGATG 480

Qy    4979 CCCTATGTGAACAAGTTCCTGCAGGATTCAAGCCCAGTGAAGAGTTGGCCAGGAGGAGC 5038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCCTATGTGAACAAGTTCCTGCAGGATTCAAGCCCAGTGAAGAGTTGGCCAGGAGGAGC 540

Qy    5039 CAGGAGAACCTCCAGGCCACAGTTAAAATATGGGGGACGCCAGCAGCTGCCGTTACCTGGT 5098
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAGGAGAACCTCCAGGCCACAGTTAAAATATGGGGGACGCCAGCAGCTGCCGTTACCTGGT 600

Qy    5099 GAAATGAATGCTTTTCTGAAA-GGGCAAGCAGTTCGTTTGCTTCTAATTCACCTGCCTGG 5157
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     601 GAAATGAATGCTTTTCTGAAAGGGCAAGCAGTTCGTTTGCTTCTAATTCACCTGCCTGG 660

Qy    5158 GGGTGTGGA 5166
           |||||||||
Db     661 GGGTGTGGA 669


RESULT 11
BG963907
LOCUS       BG963907                 825 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602928611F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4983362 5',
            mRNA sequence.
ACCESSION   BG963907
VERSION     BG963907.1  GI:14351544
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10967  Row: n  Column: 03
            High quality sequence stop: 771.
FEATURES             Location/Qualifiers
     source          1..825
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4983362"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note, this is a NCI_CGAP Library."
BASE COUNT      219 a    222 c    218 g    166 t
ORIGIN

  Query Match          10.4%;  Score 695.2;  DB 13;  Length 825;
  Best Local Similarity   91.6%;  Pred. No. 1.9e-100;
  Matches  790;  Conservative    0;  Mismatches   18;  Indels   54;  Gaps    7;

Qy    4510 CTCGGAGGCCACCTGCTTTACAACCTACAGCTCTTTGTCTGCCGACCCC-CACAACTACA 4568
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db       1 CTCGGAGGCCACCTGCTTTACAACCTACAGCTCTTTGTCTGCCGACCCCTCACAACTACA 60

Qy    4569 CC ATGCAGGAATTTGCCCTGCGCTATTTCCGGAAGCCTCATACCTGGCTGACCCAGATG 4627
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCGATGCAGGAATTTGCCCTGCGCTATTTCCGGAAGCCTCATACCTGGCTGACCCAGATG 120

Qy    4628 AGTAGAGACACCAAAGAGAAAGCTGCCATCAACCTGATCCAGTACACTAAGGACCCCATC 4687
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTAGAGACACCAAAGAGAAAGCTGCCATCAACCTGATCCAGTACACTAAGGACCCCATC 180

Qy    4688 CAGGAATCCCTTACCAGCTTCTGCAATGGGGACACAAACAGTAAAGCTGTGGCTGGCTTC 4747
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAGGAATCCCTTACCAGCTTCTGCAATGGGGACACAAACAGTAAAGCTGTGGCTGGCTTC 240

Qy    4748 AAGGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCGGGGCAAGGACGAGCTGAGT 4807
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAGGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCGGGGCAAGGACGAGCTGAGT 300

Qy    4808 CTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGACGAGATGTACTGCCAG 4867
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGGGACGAGATGTACTGCCAG 360

Qy    4868 GTCATCAAGCAAGTCACAGGACACCCCCAGCCAAAGCACTGTGCTCTGGGCTGGAGCGTC 4927
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTCATCAAGCAAGTCACAGGACACCCCCAGCCAAAGCACTGTGCTCTGGGCTGGAGCGTC 420

Qy    4928 CTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACCACGCTGATGCCCTATGTGAACAAG 4987
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CTCAGCCTCTTCACAGGCTTCTTTGCACCATCGACCACGCTGATGCCCTATGTGAACAAG 480

Qy    4988 TTCCTGCAGGATTCCAGCCCCAGTGAAGAGTTGGCCAGGAGGAGCCAGGAGAACCTCCAG 5047
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTCCTGCAGGATTCCAGCCCCAGTGAAGAGTTGGCCAGGAGGAGCCAGGAGAACCTCCAG 540

Qy    5048 CGCACAGTTAAATATGGGGGACGCCAGCAGCTGCCGTTACCTGGTGAAATGAATGCTTTT 5107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CGCACAGTTAAATATGGGGGACGCCAGCAGCTGCCGTTACCTGGTGAAATGAATGCTTTT 600

Qy    5108 CTGAAAGGGCAAGCAGTTCGTTTGCTTCTAATTCACCTG-CCTGGGGGTGTGGACTACAG 5166
           |||||||||||||||||| |||||||||| ||| ||||||||||||||||||||||||||
Db     601 CTGAAAGGGCAAGCAGTTCGTAAGCTTCTAATTCAACTGACCTGGGGGTGTGGACTACAG 660

Qy    5167 GACGAATTCACAGACATTCACAGTGGCAGGGGAAGTGCTAGAGGAGCTGTGTGGACAGAT 5226
           ||||||||||||||||||
Db     661 GACGAATTCACAGACAT---------------------------------------------- 677

Qy    5227 GGGCATCACAGACTTGGAAGAAGTGCAGGAATTTGCCCTTTTTCTCA-TGAAAGGAGAAG 5285
           |||||||  |||||||||||||||||||||| |||| |||||| |||||||||||||||
Db     678      -TCACAGAACTGGAAGAAGTGCAGGAATTTGGCCTCATTCTCATTCAAAGGAGAAG 732

Qy    5286 GTGAGCTGGTTCGGCCGCTGTACCCCATGAGTACATCAACAATGTGGTGAGGGACCAGG 5345
           |||||||||| ||| || |||||||||||||||||||||||| |||||||||||||| | |
Db     733 GTGAGCTGGTACGG-GAATGTCACCCCATGAGTACATCAA-AATTTGGTGAGGAACCAG 790

Qy    5346 ACATGAGCCTTCACAGCCGACG 5367
           |||||||||||| |||  | |
Db     791 GAATGAGCCTTCAAAGCGAAGG 812


RESULT 12
```

```
BQ924457
LOCUS       BQ924457                940 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT 3857126 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6476019 5', mRNA sequence.
ACCESSION   BQ924457
VERSION     BQ924457.1  GI:22339476
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at
            http://image.llnl.gov
            Plate: LLAM14015  row: l  column: 04
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref "taxon:10090"
                     /clone="IMAGE:6476019"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI, Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      217 a    280 c    259 g    184 t
ORIGIN

  Query Match             10.4%;  Score 653.8;  DB 14;  Length 940;
  Best Local Similarity   98.9%;  Pred. No. 1.2e-100;
  Matches  654;  Conservative    0;  Mismatches    7;  Indels    6;  Gaps    6;

Qy       5977 ATATAAAGAGCTTGGTGACCCAGGAGCTGAGGCAGATGCAAGGGTACAGCAAGCAGAGAG 6036
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 ATATAAAGAGCTTGGTGACCCAGGAGCTGAGGCAGATGCAAGGGTACAGCAAGCAGAGAG 61

Qy       6037 CATATATGGCTTTATAGAGAGCATAGCTATGTTTTTTTTTTTTTTTTACACTTTTTATG 6096
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CATATATTGGCTTTATAGAGAGCATAGCTATGTTTTTTTTTTTTTTTTACACTTTTTATG 121

Qy       6097 TATTGCTGAGAGTGAGTAAGCTGGCCCTCCCTGGACCAGGCCTCTTTGGGGTTGAACCGTC 6156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TATTGCTGAGAGTGAGTAAGCTGGCCCTCCCTGGACCAGGCCTCTTTGGGGTTGAACCGTC 181

Qy       6157 ATCTTCGTCCTCATGGACCCCAGCTCTCAGGAACTCTGCTGCTCTGTCATGTTAAAAG 6216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ATCTTCGTCCTCATGGACCCCAGCTCTCAGGAACTCTGCTGCTCTGTCATGTTAAAAG 241

Qy       6217 AAAAGCAGTTTCCATTTGTGAAGCCACTGCAGGAGGAGAAGGGCTCTGGGCTAGAAC 6276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 AAAAGCAGTTTCCATTTGTGAAGCCACTGCAGGAGGAGAAGGGCTCTGGGCTAGAAC 301

Qy       6277 TCAATATGGCTCTGTTGACAACCCTCAGACTATCTGGTTGGAGTTGCTACAGGCCAGG 6336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TCAATATGGCTCTGTTGACAACCCTCAGACTATCTGGTTGGAGTTGCTACAGGCCAGG 361

Qy       6337 AGTTTAAGAAAGATATTTCTTTGTTGGTAGCATGTTACTAGTGAGCAAGTTCT 6392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AGTTGCAGCAACCATCATGTTCCTGCTGGGCAGCATGTCCACTTACTGGCCAGGTTCT 421

Qy       6327 TCTGAGGAGTGCAGATAAGGCAGCCCTTCTCATCTGGCTAGTCTTTAATTCTCTTCT 6386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         42 TCTGAGGAGTGCAGATAAGGCAGCCCTTCTCATCTGGCTAGTCTTTAATTCTCTTCT 101

Qy       6087 GAATCCGCCTGCACAACCCCCCAACCCAAGTGCAGCCTAAAAAGTAAAGTCCTCTTCTAGTG 6146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 GAATCCGCCTGCACAACCCCCCAACCCAAGTGCAGCCTAAAAAGTAAAGTCCTCTTCTAGTG 161

Qy       6147 GGAGAATAGGCAGACACATCCCCTCTGGAGGTGGACTCTCAATATAATTTTTTCATTTTCT 6206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GGAGAATAGGCAGACACATCCCCTCTGGAGGTGGACTCTCAATATAATTTTTTCATTTTCT 221

Qy       6207 GTCTAGCCCCAGGGAAAATGCCCACCCAGCTCTGAGAGTCCTCATAATTAAAATCATCTAAGA 6266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 GTCTAGCCCCAGGGAAAATGCCCACCCAGCTCTGAGAGTCCTCATAATTAAAATCATCTAAGA 281

Qy       6267 AAAAA 6271
             |||||
Db        662 AAAGA 666


RESULT 13
BB559556
LOCUS       BB559556                638 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB559556 RIKEN full-length enriched, 2 days pregnant adult female
            ovary Mus musculus cDNA clone F430004K1 5', mRNA sequence.
ACCESSION   BB559556
VERSION     BB559556.2  GI:16449011
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouba
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sato,H., Tanaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sonobe,Y., Suzuki,H.,
            Tagami,M., Taqawa,A., Takahashi,F., Tateda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Aug 1, 2020 this sequence version replaced gi:16449009.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-exp@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            Iwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Izawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Arakawa
            ,F., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Tatada,K. and
            Hayashizaki,Y.
```

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES Location/Qualifiers
source 1..638
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E330039G21"
/clone_lib="RIKEN full-length enriched, 2 days pregnant
adult female ovary"
/sex="female"
/tissue_type="ovary"
/dev_stage="2 days pregnant adult"
/lab_host="DH10B"
/note="Site_1: SalI, Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGGGCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 144 a 195 c 179 g 120 t
ORIGIN

Query Match 10.1%; Score 634.8; DB 10; Length 638;
Best Local Similarity 99.7%; Pred. No. 5.4e-97;
Matches 636; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    5617 CTTGGTGACCCAGGAGCTGAGGCAGATGCAAGGGTACAGCAAGCAGAGAGCACAGATTGG 5676
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTTGGTGACCCAGGAGCTGAGGCAGATGCAAGGGTACAGCAAGCAAAGAGCACAGATTGG 60

Qy    5677 CTTTATAGAGAGCAAGGGCAGCTGCCCCTCTTTGGCTACACTGTGTACGTAGTGCTGAG 5736
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTTTATAGAGAGCAAGGGCAGCTGCCCCTCTTTGGCTACACTGTGTACGTAGTGCTGAG 120

Qy    5737 AGTGAGTAAGCTGGCCCTCCCTGGACCAGGCCTCCTGGGGCTGAACCGTCAGCACCTGGT 5796
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTGAGTAAGCTGGCCCTCCCTGGACCAGGCCTCCTGGGGCTGAACCGTCAGCACCTGGT 180

Qy    5797 CCTCATGGACCCCAGCTCTCAGGAACTCTGCTGCTCTGTCATGCTAAAAGACCTGAAGCA 5856
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCTCATGGACCCCAGCTCTCAGGAACTCTGCTGCTCTGTCATGCTAAAAGACCTGAAGCA 240

Qy    5857 GTTCCAGCTGCTGAGCCCACTGCAGGAGGACGGGCCCCTGGCCTAGAACTCAACTATGG 5916
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTTCCAGCTGCTGAGCCCACTGCAGGAGGACGGGCCCCTGGCCTAGAACTCAACTATGG 300

Qy    5917 CTCTGTTGACAACCCCCAGACCATCTGGTTGGAGTTGCCACAGGCCCAGGAGCTGCAGCA 5976
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCTGTTGACAACCCCCAGACCATCTGGTTGGAGTTGCCACAGGCCCAGGAGCTGCAGCA 360

Qy    5977 CACCATCATTTCCTCTCTGGGCAGCATGTCCACTCAGTGGCCAGGTCTCCTCTGAGGAGT 6036
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CACCATCATTTCCTCTCTGGGCAGCATGTCCACTCAGTGGCCAGGTCTCCTCTGAGGAGT 420

Qy    6037 GGAGATAAGGCAGCGGTCTCTTACTGGGCAGTCTGCTTTATGTCCTGCTCTGAATCCCCTG 6096
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAGATAAGGCAGCGGTCTCTTACTGGGCAGTCTGCTTTATGTCCTGCTCTGAATCCCCTG 480

Qy    6097 CACAACCCCCCACCCCACGTGGAGGCTCAAAAGGCAAAGTTGTCTCACCTGGAGAATAGG 6156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CACAACCCCCCACCCCACGTGGAGGCTCAAAAGGCAAAGTTGTCTCACCTGGAGAATAGG 540

Qy    6157 CAGACACATCCCCTCTGGGGTGGACTGCAACAGGAGTTGGGGCATTTGCTGGCTAGCCCC 6216
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAGACACATCCCCTCTGGGGTGGACTGCAACAGGAGTTGGGGCATTTGCTGGCTAGCCCC 600

Qy    6217 AGGGAAAATGCCCACCCAGCTCGAAAGCGGCACAAGTA 6254
           |||||||||||||||||||||||||||||||||||||
Db     601 AGGGAAAATGCCCACCCAGCTCGAAAGCGGCACAAGTA 638
```

RESULT 14
BF577925
LOCUS BF577925 942 bp mRNA linear EST 12-DEC-2000
DEFINITION 601091943F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4206419 5',
mRNA sequence.
ACCESSION BF577925
VERSION BF577925.1 GI:11651637
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 942)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9767 row: i column: 12
High quality sequence stop: 709.
FEATURES Location/Qualifiers
source 1..942
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4206419"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 224 a 236 c 297 g 185 t
ORIGIN

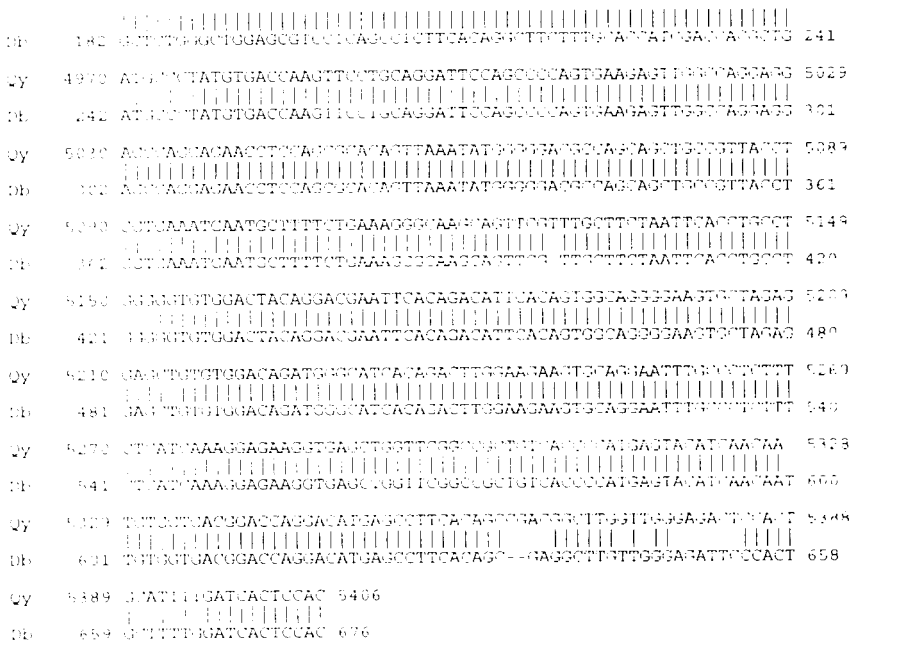
Query Match 99.0%; Score 620.8; DB 12; Length 942;
Best Local Similarity 97.6%; Pred. No. 1.1e-94;
Matches 662; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

```
Qy    4733 AAAGCTGTGGCTGGCTTCAAGGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCC 4792
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 AAAGCTGTGGCTGGCTTCAAGGCTCTGATGCAGTTTATGGGGGACCAGCCTAAGCCCCC 61

Qy    4793 GGCAAGGACGAGCTGAGTCTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGG 4852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 GGCAAGGACGAGCTGAGTCTGCTCTATGAGCTGCTGAAGCTGTGCCAAGATGACCTTAGG 121

Qy    4853 GACGAGATGTACTGCCAGGTCATCAAGCAAGTCACAGGACACCCCCAGCCAAAGCACTGT 4912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GACGAGATGTACTGCCAGGTCATCAAGCAAGTCACAGGACACCCCCAGCCAAAGCACTGT 181

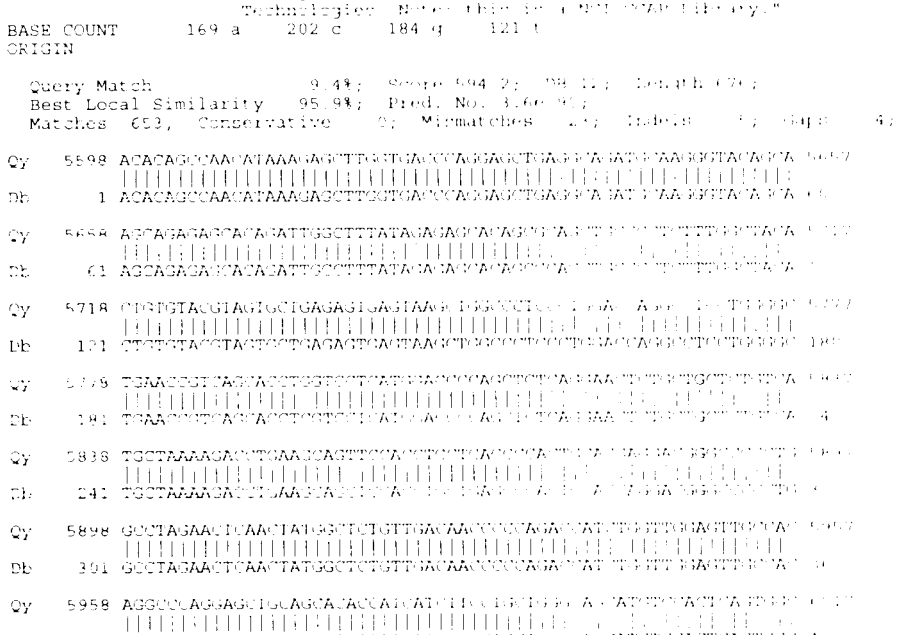
Qy    4913 GCTCTGGGCTGGAGCGTCCTCAGCCTCTTCACAGCCTCTTTGCACCAACGACCAGCTG 4969
```



RESULT 15
BF581819
LOCUS BF581819 676 bp mRNA linear EST 12-DEC-2000
DEFINITION c-2700982OF1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4219900 5', mRNA sequence.
ACCESSION BF581819
VERSION BF581819.1 GI:11956531
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 676)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbio@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM9802 row: k column: 05
High quality sequence stop: 661.
FEATURES Location/Qualifiers
source 1..676
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4219900"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.6 kb. Constructed by Life Technologies." This is a "NCI_CGAP" library."
BASE COUNT 169 a 202 c 184 g 121 t
ORIGIN

Query Match 9.4%; Score 594.2; DB 12; Length 676;
Best Local Similarity 95.9%; Pred. No. 3.6e-95;
Matches 653; Conservative 0; Mismatches 28; Indels 5; Gaps 4;



Search completed: January 10, 2003, 01:54:15
Job time : 7578.93 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

QY protein search, using sw model

Run on January [illegible], 02:[illegible]:58 ; Search time 91.3184 Seconds
(without alignments)
1938.191 Million cell updates/sec

Title US-09-803-126-1
Perfect score 6991
Sequence 1 MGSLPGPAFPGAGTPGMPTT ... WQPPSKLGPAQEVPPTEPVK 1337

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

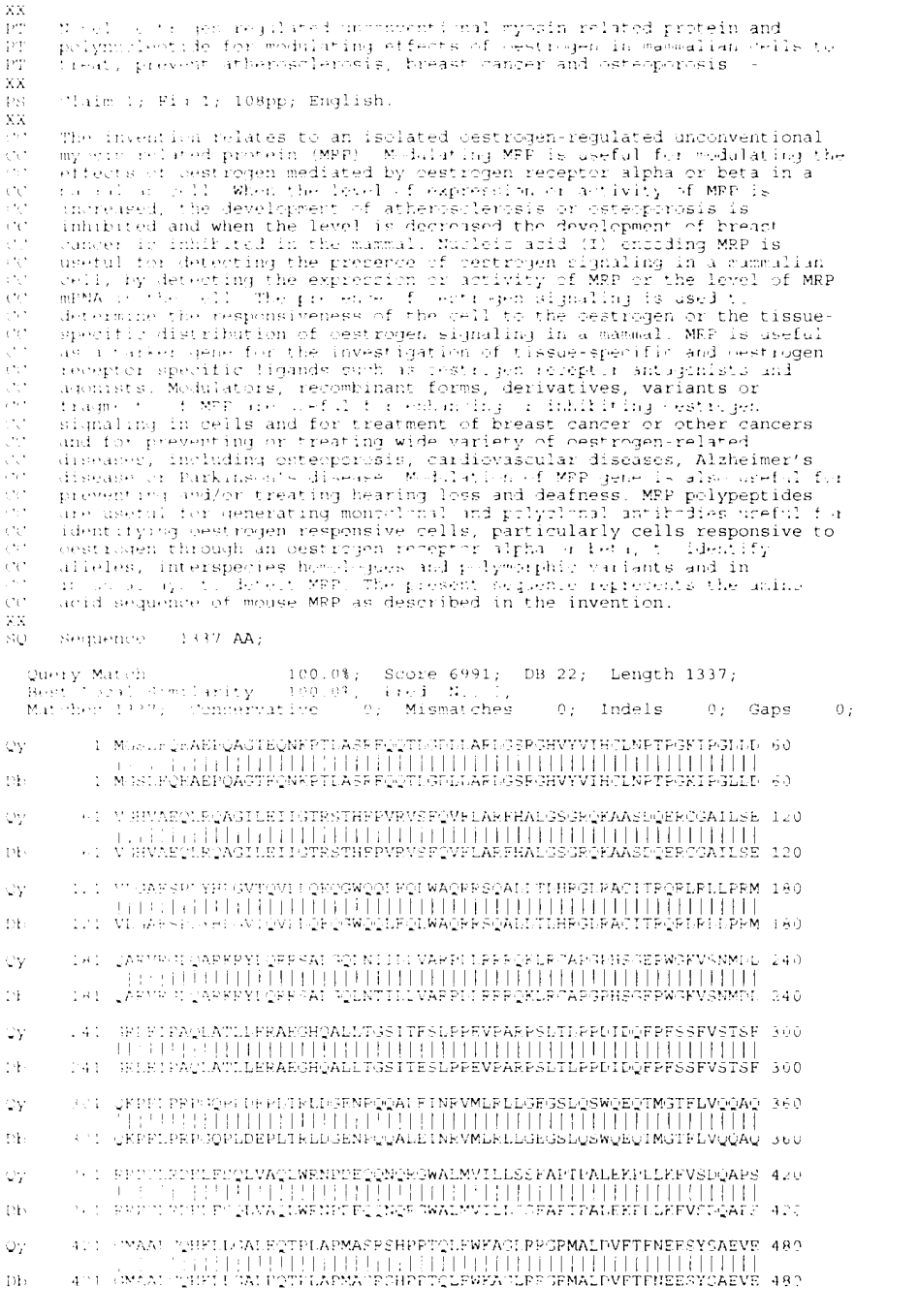
Result No.	Score	% Query Match	Length	DB	ID	Description
1	6991	100.0	1337	22	AAU07399	Mouse oestrogen-re
2	1543.5	22.1	437	22	AAU07400	Human oestrogen-re
3	1537	22.0	786	22	AAU07406	Human oestrogen-re
4	1203.5	17.2	333	22	AAU87523	Novel central nerv
5	1203.5	17.2	333	22	AAU17193	Novel signal trans
6	1203.5	17.2	337	22	AAU87214	Novel central nerv
7	1199.5	17.2	470	21	AAY57895	Human transmembran
8	920	13.2	2414	22	ABB58924	Drosophila melanog
9	895.5	12.8	235	22	AAE03412	Human gene 8 encod
10	837	12.0	193	22	AAE03410	Human secreted pro
11	443.5	6.3	188	22	AAE03409	Human secreted pro
12	328	4.7	2167	22	ABB60369	Drosophila melanog
13	315.5	4.5	572	18	AAW31855	Mycobacterium tube
14	315.5	4.5	763	18	AAW31852	Mycobacterium tube
15	310.5	4.4	446	22	ABB70063	Drosophila melanog
16	286.5	4.1	1185	20	AAY33497	Human atrophin 1 p
17	276.5	4.0	2153	22	AAU43195	Novel human secret
18	276.5	4.0	2280	22	ABB61650	Drosophila melanog
19	271.5	3.9	767	22	ABB69526	Drosophila melanog
20	271	3.9	777	23	ABP41839	Human ovarian anti
21	270.5	3.9	3119	19	AAW72204	HSV-2 strain SB5 C
22	268	3.8	1135	18	AAW31185	Human p160 polypep
23	267.5	3.8	1099	22	AAM24322	Human EST encoded
24	267.5	3.8	1404	13	AAP26049	MSF precursor. Sy
25	267.5	3.8	1404	22	AAB60568	Human megakaryocyt
26	267.5	3.8	1404	22	AAB29773	Human megakaryocyt
27	267.5	3.8	1415	22	AAU32062	Novel human secret
28	267	3.8	2063	22	ABG17147	Novel human diagno
29	267	3.8	2063	22	AAM78710	Human protein SEQ
30	267	3.8	2063	22	AAM40064	Human polypeptide
31	266.5	3.8	378	12	AAP14160	PRP 378. Triticum
32	266	3.8	2759	20	ABG17148	Novel human diagno
33	264.5	3.8	748	22	AAM40299	Human polypeptide
34	264	3.8	1012	20	AAY17406	Human atrophin-1 i
35	262.5	3.8	800	21	AAB42321	Human ORFX ORF2086
36	262.5	3.8	2703	22	ABB63299	Drosophila melanog
37	261.5	3.7	1740	21	AAY83017	Rat shank 3a. Rat
38	261	3.7	1565	22	AAM41059	Human polypeptide
39	259.5	3.7	1217	22	ABG09876	Novel human diagno
40	259	3.7	2091	21	AAB12000	Rat p3103 protein.
41	258.5	3.7	2087	22	AAM41616	Amino acid sequenc
42	258.5	3.7	4412	21	AAY83666	Sequence gi/101740
43	257.5	3.7	378	12	AAP14162	PRP encoded by clo
44	257	3.7	1414	23	AAB21728	Human PKIN-24 prot
45	256.5	3.7	1740	22	AAB21517	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAU07399
ID AAU07399 standard; Protein; 1337 AA.
XX
AC AAU07399;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mouse oestrogen-regulated unconventional myosin related protein, mMRP.
XX
KW Mouse, oestrogen regulated unconventional myosin related protein; MRP;
KW auditory; antiatherosclerotic; osteopathic; neuroprotective; nootropic;
KW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
KW Alzheimer's disease; Parkinson's disease; chromosome 11; immunogenic.
XX
OS Mus sp.
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Brooks AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS13819, AAS13820.

XX
PT [illegible] gene regulated unconventional myosin related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS Claim 1; Fig 1; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC [illegible] cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC transgenic [illegible] MRP are useful for modulating or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen-related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC [illegible]. The present sequence represents the amino
CC acid sequence of mouse MRP as described in the invention.
XX
SQ Sequence 1337 AA;

Query Match 100.0%; Score 6991; DB 22; Length 1337;
Best Local Similarity 100.0%; Pred. No. [illegible];
Matches 1337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 2
AAU07400
ID AAU07400 standard; Protein; 437 AA.
XX
AC AAU07400;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human oestrogen regulated unconventional myosin related protein [illegible] MRP #1.
XX
KW Human; oestrogen regulated unconventional myosin related protein; MRP;
KW auditory; antiatherosclerotic; osteopathic; breast cancer; [illegible];

FW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
FW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic.
XX
CC H. sapiens
XX
FH Key Location/Qualifiers
FT
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Brooks AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS13821.
XX
PT Novel oestrogen-regulated unconventional myosin-related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis -
XX
PS Claim 1; Fig 4; 108pp; English.
XX
CC The invention relates to an isolated oestrogen-regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-
CC specific distribution of oestrogen signaling in a mammal. MRP is useful
CC as a marker gene for the investigation of tissue-specific and oestrogen
CC receptor specific ligands such as oestrogen receptor antagonists and
CC agonists. Modulators, recombinant forms, derivatives, variants or
CC fragments of MRP are useful for enhancing or inhibiting oestrogen
CC signaling in cells and for treatment of breast cancer or other cancers
CC and for preventing or treating wide variety of oestrogen-related
CC diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC preventing and/or treating hearing loss and deafness. MRP polypeptides
CC are useful for generating monoclonal and polyclonal antibodies useful for
CC identifying oestrogen responsive cells, particularly cells responsive to
CC oestrogen through an oestrogen receptor alpha or beta, to identify
CC alleles, interspecies homologues and polymorphic variants and in
CC immunoassays to detect MRP. The present sequence represents the amino
CC acid sequence of human MRP variant #1 as described in the invention
XX
SQ Sequence 437 AA;

Query Match 22.1%; Score 1543.5; DB 22; Length 437;
Best Local Similarity 71.3%; Pred. No. 3.9e-94;
Matches 308; Conservative 34; Mismatches 77; Indels 13; Gaps 3;

Qy 914 MPLSPPSPGKGPPPAV 972
Db 1 MYQSPPSPVPVPQPSPPPAPLPPIDPPDEALAPLGINGAHSSPPMLSPSPGKGPPPAV 60

Qy 973 VTPPPAPPLPPLPPTPPQPPLPPTLPPQPSPMPPTAPAPPPPPALPPPLSQEPKTPSVP 1032
Db 61 APPPAPPLPLPPPPPLPPPPLPPP--QPLPIAHTPPPPPAPPLPPLPEDPGTLSAE 117

Qy 1033 SHALTPPMPPPNPSTPPLVPSPQVQPSYANAPWPLPLPKEVPYPPPNPSHPYQLSLLQQQ 1092
Db 118 RRCLTQPVPDQQVSTQLLAPSQQVQPSYTGTPWKLPLPKEVPYPRENPSHPYYLPLLCEQ 177

Qy 1093 ILRDTPTESCTRISQDERHKMKSLLGGLEVSLETLGIVPPSIPPPIVVAAPTDWAMYPGR 1152
Db 178 ILRDTPSESCIRISQNERRKMKELLGGLEVSLDSLTTIPPSVPPPIVVAAPTDWAMYPGR 237

Qy 1153 IPPVSGESGSDVQLLGVSHRGLRLLEVTQGPSPHIPQLPTLCGYSYAEVLTVQCPGPSTL 1212
Db 238 PPPVSGESGSDVQLLAVSHRGLRLLEVTQGPSLAPPQLPILCSYSFAEVIGVECRGSSTL 297

Qy 1213 ELSLKNEQLILHTAWARAIKAMVDLFLSELRKDSGYVIALAGYTTLDNSLLSPHRGDLIR 1272
Db 298 ELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALAGYTTDNSLLSPHRGDLIR 357

Qy 1273 LLPVTALEPGWQFGSAGGRSGLFPLDVVQPAAAPDLSPSLSPERNGWLP 1322
Db 358 LLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDTSPSKEQRSGWHPGQLSNGEPGLAR 417

Qy 1324 LGPAQEVRKTEE 1335
Db 418 WDRASEVRKMGE 429

RESULT 3
AAU07406
ID AAU07406 standard, Protein; 786 AA.
XX
AC AAU07406;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human oestrogen-regulated unconventional myosin-related protein KMRP #2.
XX
KW Human; oestrogen-regulated unconventional myosin-related protein; MRP;
KW auditory; antiatherosclerotic; osteopathic; neuroprotective; nootropic;
KW antiparkinsonian; cytostatic; oestrogen receptor; atherosclerosis;
KW osteoporosis; breast cancer; cardiovascular disease; deafness;
KW Alzheimer's disease; Parkinson's disease; chromosome 17q25; immunogenic.
XX
OS Homo sapiens.
XX
PN WO200168866-A2.
XX
PD 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-US08060.
XX
PR 10-MAR-2000; 2000US-188488P
PR 09-MAR-2001; 2001US-0803126.
XX
PA (SCHD) SCHERING AG.
XX
PI Brooks AR, Deng GG, Rubanyi GM;
XX
DR WPI; 2001-602747/68.
DR N-PSDB; AAS13823.
XX
PT Novel oestrogen regulated unconventional myosin-related protein and
PT polynucleotide for modulating effects of oestrogen in mammalian cells to
PT treat, prevent atherosclerosis, breast cancer and osteoporosis
XX
PS Claim 1; Fig 6; 108pp; English.
XX
CC The invention relates to an isolated oestrogen regulated unconventional
CC myosin-related protein (MRP). Modulating MRP is useful for modulating the
CC effects of oestrogen mediated by oestrogen receptor alpha or beta in a
CC mammalian cell. When the level of expression or activity of MRP is
CC increased, the development of atherosclerosis or osteoporosis is
CC inhibited and when the level is decreased the development of breast
CC cancer is inhibited in the mammal. Nucleic acid (I) encoding MRP is
CC useful for detecting the presence of oestrogen signaling in a mammalian
CC cell, by detecting the expression or activity of MRP or the level of MRP
CC mRNA in the cell. The presence of oestrogen signaling is used to
CC determine the responsiveness of the cell to the oestrogen or the tissue-

```
CC   specific distribution of oestrogen signaling in a mammal. MRP is useful
CC   as a marker gene for the investigation of tissue-specific and oestrogen
CC   receptor-specific ligands such as oestrogen receptor antagonists and
CC   agonists. Modulators, recombinant forms, derivatives, variants or
CC   fragments of MRP are useful for enhancing or inhibiting oestrogen
CC   signaling in cells and for treatment of breast cancer or other cancers
CC   and for preventing or treating wide variety of oestrogen-related
CC   diseases, including osteoporosis, cardiovascular diseases, Alzheimer's
CC   disease or Parkinson's disease. Modulation of MRP gene is also useful for
CC   preventing and/or treating hearing loss and deafness. MRP polypeptides
CC   are useful for generating monoclonal and polyclonal antibodies useful for
CC   identifying oestrogen responsive cells, particularly cells responsive to
CC   oestrogen through an oestrogen receptor alpha or beta, to identify
CC   alleles, interspecies homologues and polymorphic variants and in
CC   immunoassays to detect MRP. The present sequence represents the amino
CC   acid sequence of human MRP variant #7 as described in the invention
XX
SQ   Sequence   786 AA;

  Query Match             21.2%;  Score 112;  DB ..;  Length 786;
  Best Local Similarity   24.0%;  Pred. No. 2.2e-93;
  Matches  83;  Conservative   32;  Mismatches   70;  Indels    4;  Gaps    2;

Qy      911 MT[illegible]VPIPAFLPPNIPIPIPALAPLTINTVH ITLPTTRIQPSPPFAV 970
            ||||||| |||||:|| | | :| ||||||||||||||| | |  ||: ||  ||||
Db        1 MTQMPSPVPVPVQLPPPAFLPPIIPIPIEALAPLSINGAHSSIPMLIPSPGPGPPFAV 60

Qy      971 VPPTPAAPPLPLPPGLPIQPPQGPLPLLPGPGSPNPPTAPAPPPPPALPPPLSGPPETPSVE 1030
            |  ::   :| |: ||:|||||| |||;      | |  |||||| | ||  :| | | |
Db       61 APKPPAPIQLGPSSSIPPPQGPIIIPG---QSIPIAHTPPPPPAPPPLPPPPGTISAP 117

Qy     1031 SHALTEPMEDYNICTFLLVPCSCVCFSVANAFWFIFFPEVPYTPFENFCHPYTLSLL 1090
            :  ;;:|: :|:|:|  |||||||||||||||||||||||||| | |||:|
Db      118 RPCLTQIVEDQGVSTQLLAPSSSVLPSYPSTPWPIPIPPEVPYTPPENPSHPYYLSLL 177

Qy     1091 IIPPTTTPLPIPISQPEPPPMPSIISPLPVSIPILPIVEPSIPPPIVVAAPFNWANYPSP 1150
            |||:|:|| |||:|| ||| ||| ||| |::|   |||:||||||||||||||||||||
Db      178 IISIPPSPSCIPISQNEPPPMPDILSSLPVLLQLTTTEECVPPPIVVAAPFNWANYPSP 237

Qy     1151 IPPVSQPPQSDVQLLGVSHPGLPIIPVTQSPSPHIPLPFIPPPLIAPVITVLPPPPIL 1210
            :  : ||:||| |||:|||||||| ;    |||| |||||:|||| |:||| |||
Db      238 FPPVSCPSQSDVQLLAVSHPGIPIIPVTQGPGIPPIQIPIIPGSYSPAEVIGVEPGGSTL 297

Qy     1211 ELTLRNFLQIIHTAWAPAIPAMVDLPLDELPPDSGYVIALPSTITDDNSLLSFHPGDLIP 1270
            ::::: ||:|||| ||||:|:|:||:||:||||||||||||||: ||||||||||:
Db      298 PIQIPSTLQVLHTAPAGAIPALVELPLNELPPIIGYVIALPSTITDDSSLLSFHPGDLIP 357

Qy     1271 DIPA AIPPPWLPPSAPPPPPVPAAAPIISPPLGPPNSWIP 1330
            ;:    :|||||||||||:|||| |:|||||||| ||| :|: | :
Db      358 DIPVNPLPPPGWQPQSAQSPPVLPPADIVLPAAAHTPSPSPPLPPSGWHP 406


RESULT 4
AAU87523
ID   AAU87523 standard; Protein; 333 AA
XX
AC   AAU87523;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Novel central nervous system protein #433.
XX
KW   central nervous system; CNS; autoimmune disease, rheumatoid arthritis,
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder, Alzheimer's disease, AIDS, ocular disorder,
KW   acquired immunodeficiency virus, dysphagia, gastrointestinal disorder,
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer, leukaemia, neovascularisation,
KW   respiratory disorder, renal disorder, kidney failure; blood disorder,
KW   [illegible] activation, wound healing, cell proliferation, skin aging,
KW   food additive; food preservative; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200155318-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001332.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0190081.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0206515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   [illegible]; [illegible]
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218230.
PR   26-JUL-2000; 2000US-0220963.
PR   04-AUG-2000; 2000US-0223044.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231240.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234797.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
```

PR 29-SEP-2000; 2000US-0236834.
PR 29-SEP-2000; 2000US-0236836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241810.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249268.
PR 17-NOV-2000; 2000US-0249269.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250201.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251388.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251930.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0253628.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM,
XX
DR WPI, 2001-581633/65
DR N-PSDB; ABK43853.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 9; SEQ ID NO 1041; 837pp; English
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 17.2%; Score 1283.5; DB 22; Length 333;
Best Local Similarity 72.8%; Pred. No. 9.7e-72;
Matches 236; Conservative 29; Mismatches 50; Indels 9; Gaps 1;

Qy 1021 PLSGEPKTPSVESHALTEPMEDKNISTKLLVPSGSVCFSYANAPWKLFLRKEVFYPKENF 1080
Db 2 PLPSFQPGTLSAEFRCLTQPVEDQGVSTQLLAPSGSVCFSYTNTPWFLFLRKEVFYPRENF 61

Qy 1081 SHFYCLSLLCQQILSDTFTPSCTPISQDPPHPMKGLIGDLEVSLETLDIVEDSIKKRIVV 1140
Db 62 SHPYYLSLLCPQILPDTFSESCTPISQNEPFPMKTLIGDLEVSLETTPTSNFKRIVV 121

Qy 1141 AAEDNWANYFSRFFPVSGESGSDVQLLAVCHPGLPLLPVTQHPSFHLDQLFTLGSTGTAE 1200
Db 122 AAEDNWANYFSRFFPVSGESGSDVQLLAVCHPGLPLLPVTQHPSFHLDQLFTLGSTGTAE 181

Qy 1201 VLTVQNPGPSTLELSLKNEQLILHTAWAPAIKAMVDLFLSELPKDSGYVIALRSYITDDN 1260
Db 182 VLGVEIPGGSTLPGSLFSFLLVTHTAPAPAIPALVELFLNELVPDSGYVIALPSYITDNC 241

Qy 1261 SLLSFHFGDLIPLLPVTALELQWQFGSAPGPSGLFPDLVVQPAAAPDLSPSLGPPNPWQP 1320
Db 242 SLISFHPGDLIPLLPVATLELQWQPGSAGGPSGLFPAIVVQPAAAPDPSPSPQPSGWHP 301

Qy 1321 ------FSKLGPAQEVPKTEE 1335
Db 302 SQLSHIPEGLAPWDPASEVPPMGE 325

RESULT 5
AAU17193
ID AAU17193 standard; Protein; 333 AA.
XX
AC AAU17193;
XX
DT 07-NOV-2001 (first entry)
XX

DE Novel signal transduction pathway protein, Seq ID 758.
XX
KW Neuroprotective, cytostatic, dermatological; immunosuppressive; tumour;
KW anti-inflammatory, anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis, inflammatory condition;
KW renal transplant rejection, infection, hepatitis C, blood disorder;
KW [illegible] cell anaemia, hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder, Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome, ischaemia; renal disorder;
KW cardiovascular, respiratory, wound healing, endocrine, Addison's disease;
KW reproductive system, gastrointestinal; liver disorder; AIDS;
KW Acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO[illegible]733 A1.
XX
PD [illegible] 2001.
XX
PF 17-JAN-2001; 2001WO-US01112.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 04-FEB-2000; 2000US-0180664.
PR 06-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 10-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 08-JUN-2000; 2000US-0210886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224571.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227102.
PR 24-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 13-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239936.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251856.
PR 05-DEC-2000; 2000US-0251868.
PR 05-DEC-2000; 2000US-0251869.
PR 05-DEC-2000; 2000US-0251989.
PR 05-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Parish SC, Ruben SM;
XX
DR WPI; 2001-465160/50.
DR N-PSDB; AAS27110.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID NO 758; 980pp; English
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 17.2%; Score 1203.5; DB 22; Length 333;
Best Local Similarity 72.8%; Pred. No. 9.7e-72;
Matches 236; Conservative 39; Mismatches 50; Indels 9; Gaps 1;

Qy 1321 -- ------KSKLGPAQEVPKTEE 1335
:: | |||| |
Db 302 GQLSNGEDGLARWDRASEVRKMGE 325

RESULT 6
AAU87214
ID AAU87214 standard; Protein; 337 AA.
XX
AC AAU87214;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #124.
XX
KW Central nervous system, CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder, neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder, ischaemia, angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder, testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder, renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01322.
XX
PR 31-JAN-2000; 2000US-0179065
PR 04-FEB-2000; 2000US-0180628
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009
PR 30-AUG-2000; 2000US-0228924
PR 01-SEP-2000; 2000US-0229287.
PP 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

PR 07-SEP-2000; 2000US-0229345.
PR 07-SEP-2000; 2000US-0229509.
PR 07-SEP-2000; 2000US-0229513.
PR 07-SEP-2000; 2000US-0229437.
PR 07-SEP-2000; 2000US-0230428.
PR 08-SEP-2000; 2000US-0231042.
PR 08-SEP-2000; 2000US-0231043.
PR 08-SEP-2000; 2000US-0231044.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 13-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232497.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234273.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251068.
PR 05-DEC-2000; 2000US-0250719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-581633/65.
DR N-PSDB; ABK43544.
XX
PT New isolated nucleic acid encoding a protein for treating,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives
XX
PS Claim 9; SEQ ID No 732; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I), and fragments, vectors, host cells
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysplasia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level, e.g.
CC leukaemia, disorders involving peripheral nervous system malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 17.2%; Score 1203.5; DB 2; Length 837;
Best Local Similarity 72.9%; Pred. No. 9.9e-72;
Matches 236; Conservative 29; Mismatches 57; Indels 9; Gaps 1;

Qy 1021 PLSGFPFTPSVESHALTEPMEDRNISTKLLVPSGSVENYANAEWPLPIFSPPVPTIEEQP 1080
|| :| | | | |:|:|| :|:| | |||||||| ||| | | |
Db 6 PLPEDPGTISAFPPCITQPVPTQVVSTQTIAPSQVVSSVTSTTWPTSSEPVTTIEENP 65

Qy 1081 SNPSISLSQWFFKFDFSTISQFEEHMMCCLHRVCKECMDYCSPPSYV 1134
|||| | |||||||||||:||| |||:|| || ||| ||| :|:| |||||||
Db 66 SHPYYILKLLCEQILKDIPSESCIPISQNEKKAMAGLGCLEVCCCTTTEDSVEEFIVV 125

```
Qy    1141 AARDNWANYFSRIFFVSGESGSDVQLLGVSHPGLPLLKVTQSPSFHLDQLKTLCSYSYAE 1200
           ||||||||||||| |||||||||||||| |||||||||||| |    |||| |||||:||
Db     180 AARDNWANYFSRFFFVSGESGSDVQLLAVSHPGLPLLKVTQGPGLPPDQLKTLCSYSFAE 239

Qy    1201 VLTVQCRGRSTLELSLKNFQLILHTAWAPAIKAMVDLPLSELRKDSGYVIALRSYITDDN 1260
           || |:||| |||||||:|||:|||| ||||:|:|:|||:||:|||||||||||||||||:
Db     240 VLGVECRGGSTLELSLKSEQLVLHTAPAPAIEALVELPLNELKKDSGYVIALRSYITDNC 299

Qy    1261 SLLSFHRGDLIKLLPVTALEPG--W----QPGSAGPSPSGLPPDVVQPAA 1304
           |||||||||:|||  ||   |      :  :|| || ||   ::| |
Db     300 SLLSFHRGDLIKLLPV--CHPGARLAVWLCRGPPPTLSCPHSAAGPSSLP---LQQGAE 354

Qy    1305 APDLSFSIGKPNSWQRKSKLGPAQEVRKTEE 1335
                 | : :: || :: |   :  : ||
Db     355 E-----WLAQGSAVQRGTPAGSVGQGLPGPE 380
```

RESULT 7
AAY57895
ID AAY57895 standard; Protein; 470 AA.
XX
AC AAY57895;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN 19
XX
KW Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11904.
XX
PR 29-MAY-1998; 98US-0087760.
PR 02-JUL-1998; 98US-0091674
PR 02-OCT-1998; 98US-0102354.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
PI Au-Young J;
XX
DR WPI; 2000-072605/06.
DR N-PSDB; AAZ56716.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders -
XX
PS Claim 1; Page 125-126; 229pp; English.
XX
CC AAZ56699 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or
CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.
XX
SQ Sequence 470 AA;

  Query Match 17.2%; Score 1199.5; DB 21; Length 470;
  Best Local Similarity 64.5%; Pred. No. 2.8e-71;
  Matches 252; Conservative 37; Mismatches 73; Indels 29; Gaps 6;

```
Qy     961 SPNQGPSSPPAVVPPPKAPPPLEPSLSIQEKQGPLPDLFGPQSPNPPTAPAPPPPPALPP 1020
           ||: ||  |||| |||||  :| || ||:|||||| ||||       | |  |||||| |
Db       3 SPSPGKGPPPAVAPPPKADLQLGPSSSIKEKQGPLLDLFG---QKLPIAHTPPPPPAPPL 59

Qy    1021 PLSGEPKTPSVESHALTEPMEDKNISTKLLVPSGSVGPSYANAPWKPLPKPVPVPPPNP 1080
           ||  :| | | |   ||:|:||: :||:|| |||||||||   ||||||||||||||||
Db      60 PLPEDPGTLSAEPPQLTQIVEDQSVSTQLLAPSGSVGPSYTTTPWPLPLPKPVPVPPPNP 119

Qy    1081 SHPYCLSLLCQQILRDTFTESCTPISQDEPPKMKGLLGDLEVSLETLDIVEDSIKKPIVV 1140
           |||| | |||:|||||||:||| ||||:|| ||| ||| ||| |:|    |||:||||||
Db     120 SHPYYLRLLCEQILRDTFGESCIPISQNEPPPKMKDLLGGLEVDLDSLTTTEDSVKKPIVV 179

Qy    1141 AARDNWANYFSRIFFVSGESGSDVQLLGVSHPGLPLLKVTQSPSFHLDQLKTLCSYSYAE 1200
           |||||||||||| ||||||||||||| ||||||||||||| |    |||| |||||:||
Db     180 AARDNWANYFSRFFFVSGESGSDVQLLAVSHPGLPLLKVTQGPGLPPDQLKTLCSYSFAE 239

Qy    1201 VLTVQCRGRSTLELSLKNFQLILHTAWAPAIKAMVDLPLSELRKDSGYVIALRSYITDDN 1260
           || |:||| |||||||:|||:|||| ||||:|:|:|||:||:|||||||||||||||||:
Db     240 VLGVECRGGSTLELSLKSEQLVLHTAPAPAIEALVELPLNELKKDSGYVIALRSYITDNC 299

Qy    1261 SLLSFHRGDLIKLLPVTALEPG    --W      ----  QPGSAGPSPSGLPPDVVQPAA 1304
           |||||||||:|||  ||   |      :  :|| || ||   ::| |
Db     300 SLLSFHRGDLIKLLPV--CHPGARLAVWLCRGPPPTLSCPHSAAGPSSLP---LQQGAE 354

Qy    1305 APDLSFSIGKPNSWQRKSKLGPAQEVRKTEE 1335
                 | : :: || :: |   :  : ||
Db     355 E-----WLAQGSAVQRGTPAGSVGQGLPGPE 380
```

RESULT 8
ABB58924
ID ABB58924 standard; Protein; 2424 AA.
XX
AC ABB58924;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 3564.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL03027.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 3564; 21pp + Sequence Listing; English.
XX

```
CC  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology, and in elucidating cell signalling and
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (ABL16175-ABL16179, expressed DNA
CC  sequences (ABL01840-ABL16175) and the encoded proteins
CC  (ABL00001-ABL00671).
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence    2424 AA;

  Query Match             13.2%;  Score 920;  DB 22;  Length 2424;
  Best Local Similarity   25.0%;  Pred. No. 8.2e-62;
  Matches  814;  Conservative  320;  Mismatches  560;  Indels  586;  Gaps   61;

Qy      19 PTLASFFQQTLGDLLARLGSPGHVVYVIHCLNPTPGKIPGLLDVCHVAEQLRQAGILEIIG 78
           .: :|   :|  || :| | |  : :: |   |   :|:  | :||| |:|: |
Db     763 PTVAAFFADSLQQLLQSMGPCHPWFVPCIFPNQEKHALPMDMPCVLQQLPYLGMLDTIQ 764

Qy      79 TSPTHPFVPVSPQVFLAPF HALGGQPQPAAGLQEPCCAILSFV  GAFSPLYHLGVTQ 135
           :::|: :| | |: :  :    :| | | :|  :   | | | | || |:
Db     765 IFLPGYFVPLPFQHPVEPIPHLIPSPIAPAIIPIPPIQPAILEAMPFTPVEGPGYQLGATP 824

Qy     136 VLLQPQGWQQLEQLWAGPPSQALLQIMPGLPACITPQPL ---PILPPMQAPVPGLQAPK 191
           : | :|   :| :|:: | :|  : |::|       |:||| || :|..
Db     825 VPLPEALPEALESGRTEPLPPAAVSVPPHVPLMLVPPQLAPPQAAATPLQAPWPGQPAPQ 884

Qy     192 PYLQ PPSAL  ---  -GQLNTILLVAPPLLPPPQKLPCAPGPHSCEPWGK---VSNMD 239
           :: : :| |      |    :  |  ||||: | |             |   :
Db     885 PYPPLSPGALTAQPLWPGPQARRPVLLPSLHRPPQEAPEAAQPAPPAPPAPQAVLPPSQ 944

Qy     240 LGPLEIPAQLATLLEPAFG--------HQALLTGSITESLPPEVPAPPS--LTLPPPIDQ 289
           :  |:| :  : :|        |   : |:          || ||  : || |: |
Db     945 LSYLDIPAPLAPITSALQGWSFPHGGRHLVRVLGI      - VPGIPSGAVQLPEDLGQ 996

Qy     290 PPPSPPVSTPPQKPPLPPPQGPLDFP-LTPIPGFNP--QQALPINPVMLPLLGEGSLQSW 346
           |  | |        |   :|:  | |||     :   | || : :::||   : :|:
Db     997 PSPLPDLGVYQNGLPLQPPEPITAPLLTPAASPPQPPQPALAVPPLILPWSNQVALEGA 1056

Qy     347 QELTMHIPLVQQAQPPPQLPTDFIPSQLVAQL-WPNPIPQQNQPGWALMVILLSGPAPTPA 405
           : : :  :: | :|    |||||:  || |:    :  | | |: |  | |:| |
Db    1057 PPLLAPLIVPKPALSSPSLPDPLLVPLCNQVPGLPPNSGPATPIWQIIGGQTLCCPQPSAA 1116

Qy     406 LEKPLLPPVSDQAPSCMAALCQHKLI----CALPQTPLAPMASPSHPPTQLEWKAGLPPG 461
           :  ::|| |:||   :|   :||     |       :  | ||  |  |||:|  |
Db    1117 PPPYLMKPVDDEAPESLRPLLLPQLLRQQGGGTSSGAVGAGACPSFVPAWLPWPAWTPGC 1176

Qy     462 PMALPVPTPNKESYLAEVESWIIGPQPAGWIIQSPGLPAPPPGWPVSLHSGDAWPDLPGC 521
           ||  :   :| | :  |:||: |: |    : | |: :  |||:: |  |     |  |
Db    1177 PPLLPVTLPPLAGQLPVAVLPAPLIPPAAAIAVPGIPVAG  PGWIIVILPQQLTEPPL 1234

Qy     522 DPVLDLIGQTP      ----  ------- --DLGD----PAGPHNYPITPLGLAESIP 555
           |:|:|  : |                    | |    || |   |  |  |  | :|
Db    1235 IPPMLITAEPELIPAPTAPPGIIPPGAPPAPTTLPDAVPPPAVPPPAPPTSPG-PPDVP 1293

Qy     556          -    -- -------PAPG--  ---------------------VQAPS-- 564
                          |:||                              | |:
Db    1294 RPIPSPRELIPPSSALNEPYPEPEPSPSPGPQGSSTGQAPSPSPSLPDILAGDIVPVPTDC 1353

Qy     565          -    ---  --------LPPGLPPGP------------------------ 573
                             | |   ||
Db    1354 PPPLPDLGLSESKLNDRIHSAERIAPVGRLTGPPYQPSQHAIPPSHAASHGSHSSKYA 1413

Qy     574 -    -    -   -APILASS  --RPPGEASKPENLDGFVDHLPEPALAPGFSDLE 612
                        || ;||     | | :|   ::| |:| || | |     :|
Db    1414 PKAPYATRPVAMPQTTGPAPGIAPPVPPVPPQAS  PVPLPPLPLLPSPVIPPSPLPPLS 1473

Qy     613 QGAAIGIMPGLGD       -- --------DVGPTQQGYPMVY----------------P 635
           :: : ::||                    :  |  |:|
Db    1474 PASPLAAPIPGSPVPLPLPEAPLPPPLLVIPNPIPPPIPLVPNLPPLTPVPLPAAIIPPP 1533
```

```
Qy     640 GMVQAPSTQPA              MIPAPMPVMPAM --                    GAV 679
           |:   |  :                  :  |:||  :                     ||
Db    1534 GVALPPLPPEEE.PIPLPLPLVPNPPPPPPPP :PPP PPPPPPPPPPPPPPPPPPPPAP 1593

Qy     680 PTMPAMMVPPPPQPPVPPPPPPQL        APPPPVPPPPLLPVPLPPMTTQAMPLPLE 739
           | :  :|  |      |:|  :||         |:  |  ||  :|||    |:|  |
Db    1594 PLLPQLPMQLPPSPTPPSPPALPVPPPPPPPPAPPLPPAMAPNPAPPPP  PPPPP  PQAPPPPLPP 1643

Qy     740 PPPPPPPPPPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 799
           ||        |  : :|| || |  :  :  || ||           | :|:|      |
Db    1644 QQ-----AAAAAANNSTSPPPAPPPPSSPSPSPPPPPQPPMPMPPATPPPPVPPSPTPHPPPP 1698

Qy     800 PEEAES-------  -KPQPPKSPQQKKDYPQ KMSQ                        - 842
               |::          ||    || : |:|  || |
Db    1699 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1758

Qy     843 ------------DPIRVKTVKPP   AKVQIPQEEMEET           EEEEVETAEL 884
                       :  :::  ||   ::      :| |             : : || ::
Db    1759 PGARGASATAIPWPPPPIPSPTPPPPMSQQMPSQIRFTPTPTPTPTPTAVPNKPKPKPPQPV 1818

Qy     885 SPP----------  -        --PPPPPVVKKPLKASPPK                AVPEI 924
             |                      |||  | |  |  :   :                 : | :
Db    1819 GQPMQSVVTTKLAPKSPPIQALPPPPMQVQKLPISSPMRQPIPQVTAHPVPNTVSTPPSP 1878

Qy     925 EAEPAQEE---  -  ---  -        --          -          VITQGR        IEIPVH 972
           :  ||: |                                              | :|          ||
Db    1879 QRAPAKLEIPKKIMLQQQLSQLIPAQVPQPNQVPTPVPQPHPLSPQPQPMPPPPPPAP 1938

Qy     973 SSNS--------  APQHPKPSPVPPVQSSNSAPPRPQPSREIPNIIPMYQPPGPVAVPVQ 1026
           |             ||  | |    ||::  : |||  | |: :      :  :  |
Db    1939 PSGGWPPVLLPPAPSVPAPP--PPIPPPSMAPPAPPPAPQNIITAPQIPEPEPDYETSSSQ 1996

Qy     927 PTR---PIKTPQKKNDPKDEALAKLGINGVHLPLSTSPNQPPLTIIAVV             979
           :  |    :: |          :  |:  || |    :  :| :||
Db    2000 VVKEHVPAPIQPQERDTPGAVPQQQMISSHHLIPLE   IHPAQSSPAVAAWPQAEPPPGP 2059

Qy     980 PPPKAPPRLEPSISI---QPKQGPL    ---          -          -  RILPPIP 1031
           | :| | :  | |:     | :||                              ::
Db    2060 QRGRGPIPEDYGELVWPPAPVPPPAQQQQQPPPPPPPPPPPPPPLVPIPQPPPPPPPPPP 2119

Qy     1032 SPNPPTA -----  ----------  -  -  -                          - 1059
            | |
Db    2120 QPAPIPPVPQASMPITGGHPPPIPTGPIAIPPPTPMAQPYPHPPPPPPPPPPPPPMPPPPLLQMPH 2179

Qy     1060 -PAPPP---PPALPPPLSGPIKTPSVPESHALTPIMPDPNIGTPLLVPPSP    VPPPYA 1112
              |||    |  :||| |  |                 : :: : |  : :|
Db    2177 MTTPPPVIVPAPVPPPPSSTSP --      ---      GLGVVPPVLMGPPGGSTAPLTYN 2231

Qy     1062 NAPWKIPIPKKEVPYPPPENPGPPYPLSIPPQQTIPLTPAPPPPPPPPPPPPPPPPPPP 1121
             ||| :|||| | ,    |  | ||   |: | |  |      | ||:  :|  :|
Db    2232 PVPWPLPVPPPEVPPPHPPPIPPPVAIPIPPAQVPPPVPVTPPPPPPPPPPPPPPPPPP 2291

Qy     1122 EVSLETLD -          -      IVELPIKKPIVVAAPINWANYPLPIPPVPPPP  PPPVQ 1174
              ||:||                |   :|: :|  ||:    ,,::: | :  |      ||
Db    2292 GVSVPTLPNPNGGPAGGGGPNGPQVPPALVPPPHLVPMAPP WPLYPPAPLPAVQPAPIYPPVP 2351

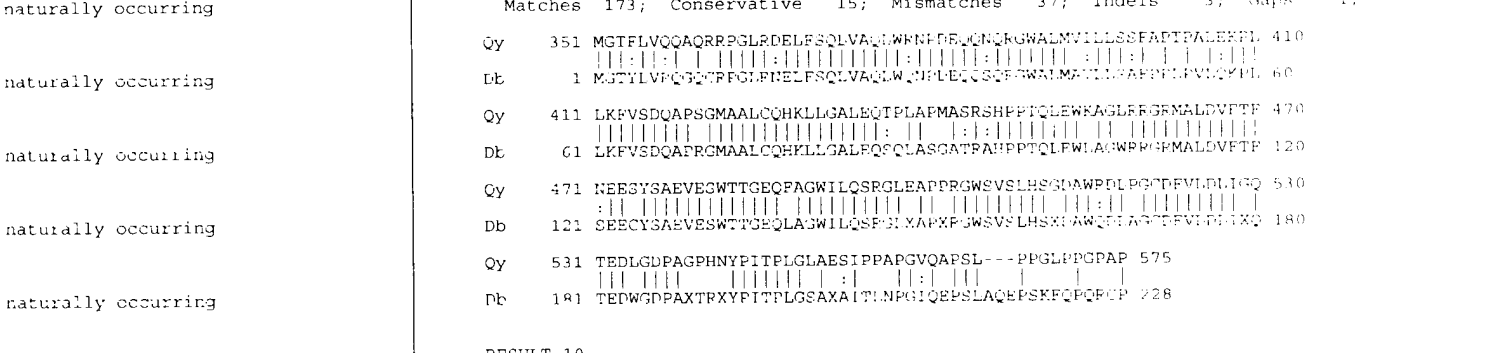
Qy     1175 LLGVSHPPGLPLLPVPQSPSPHLPQLPPTLPSYSYPAPVLTVQ P PP PTLPLPIPPN PQLII 1234
            ::||| || | :           | |  : : |: | :          :  : : | : | :
Db    2352 IMGVSHSGLYLAP --- PPAPYLIVVQAISPPPPIQPAVTIPPPAAPLPLNPPNPPPPPAL 2404

Qy     1224 HTAWARAIKAMVLLFLSELRK 1244
            | | | ||::||   |: | ||
Db    2405 HAARAAAIQSMVTTFVQEPPK 2425
```

```
RESULT 9
AAE03412
ID   AAE03412 standard; Protein; 235 AA.
XX
AC   AAE03412;
XX
DT   1-[illegible]-2001  (first entry)
```

XX
DE Human gene 8 encoded secreted protein variant. SEQ ID NO:70.
XX
KW Human; secreted protein, proliferative disorder, cancer, tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease, cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerary;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 149
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 152
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 162
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 179
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 189
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 192
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 202
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 232
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 234
FT /label= Unknown
FT /note= "Xaa equals any of the naturally occurring
FT L-amino acids"
XX
PN WO200132687-A1.
XX
PD 10-MAY-2001.
XX
PF 29-OCT-2000; 2000WO-US29364
XX
PR 29-OCT-1999; 99US-0162738.
PR 30-JUN-2000; 2000US-0215134
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Duan RD, Fiscella M, Ebner R,
XX
DR WPI; 2001-308740/32.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
FT used in prognostic, treating or ameliorating a medical condition -
XX
PS Disclosure; Page 22; 398pp; English.
XX
CC AAD07770-AAD07794 represent cDNAs corresponding to 10 human secreted
CC protein genes, and AAE03378-AAE03402 represent the proteins they encode.
CC AAE03403-AAE03415 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 10 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein variant referred to in the disclosure of the invention.
XX
SQ Sequence 235 AA;

Query Match 12.0%; Score 895.5; DB 22; Length 235;
Best Local Similarity 75.9%; Pred. No. 1.7e-51;
Matches 173; Conservative 15; Mismatches 37; Indels 3; Gaps 1;



RESULT 10
AAE03410
ID AAE03410 standard; Protein; 193 AA.
XX
AC AAE03410;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human secreted protein fragment, SEQ ID NO:68.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease, cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;

```
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc_difference 160
FT                   /label= Unknown
FT                   /note= "Xaa equals any of the naturally occurring
FT                   L-amino acids"
FT   Misc_difference 163
FT                   /label= Unknown
FT                   /note= "Xaa equals any of the naturally occurring
FT                   L-amino acids"
FT   Misc_difference 173
FT                   /label= Unknown
FT                   /note= "Xaa equals any of the naturally occurring
FT                   L-amino acids"
FT   Misc_difference 190
FT                   /label= Unknown
FT                   /note= "Xaa equals any of the naturally occurring
FT                   L-amino acids"
XX
PN   WO200132687-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US29364.
XX
PR   29-OCT-1999;   99US-0162238.
PR   30-JUN-2000; 2000US-0215134.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Komatsoulis GA, Duan RD, Fiscella M, Ebner R;
XX
DR   WPI; 2001-308740/32.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition
XX
PS   Disclosure; Page 383-384, 398pp; English.
XX
CC   AAB07770-AAB07794 represent cDNAs corresponding to 19 human secreted
CC   protein genes, and AAE03378-AAE03402 represent the proteins they encode.
CC   AAE03403-AAE03415 represent human secreted protein fragments or variants
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   19 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein fragment referred to in the disclosure of the invention.
XX
SQ   Sequence   193 AA;

  Query Match             12.0%;  Score 847;  DB 2;  Length 193;
  Best Local Similarity   81.3%;  Pred. No. 9.8e-48;
  Matches  152;  Conservative   15;  Mismatches   21;  Indels    5;  Gaps    5;

Qy     340 EGSLQSWQSQTMGTFLVQQAQRRFGLRDELPQQLVAQLWENTPLLRQF 387
           :|||:||| | |||:|||:| | |||||:|||||||||:|:||| ||:|| |||||| :|||:
Db       1 MPMFPSWQRQIMGTYLVRQSQCRFGLRNELPQQLVAQLWQNTPFLLQF 48

Qy     388 FAFTFALEKPLLKTVSDLAFSPMAALCQHFLSNIFYTPLAPMAVGSHTDPTQLPWKACLP 447
           | | | |:|||:|||||||| ||||||||||||||:|:|| :|  :|:|  |:|:| |
Db      49 FPPLPVLQFPLLPFVTDLAFPPMAALCQHFLSAIFVTPLAPLAVGSQTDPTQLPWTAWSA 108

Qy     448 RGRMALDVFTFNEESYSAEVESWTTGRQFAGWTLGGLFAAFPAWWVLHFGTAWPLLP 504
           |||||||||||:|| |||||:|||| |||||:|||| |:|:|:|:|:|:|:|:|:|:
Db     109 PGPMALDVPTPTPPCYSAFVESWTTGRQLAGWTLGGLFAAFPSWWVLHFGSDAWQDLA 166

Qy     505 GGDFVLDLIGQTE 532
           ||||||||| |||
Db     167 GGDFVIDLIKQTF 193


RESULT 11
AAE03409
ID   AAE03409 standard; Protein; 188 AA.
XX
AC   AAE03409;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein fragment, SEQ ID NO:67.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
PN   WO200132687-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US29364.
XX
PR   29-OCT-1999;   99US-0162238.
PR   30-JUN-2000; 2000US-0215134.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Komatsoulis GA, Duan RD, Fiscella M, Ebner R;
XX
DR   WPI; 2001-308740/32.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition
XX
PS   Disclosure; Page 382-384; 398pp; English.
XX
CC   AAB07770-AAB07794 represent cDNAs corresponding to 19 human secreted
CC   protein genes, and AAE03378-AAE03402 represent the proteins they encode.
CC   AAE03403-AAE03415 represent human secreted protein fragments or variants
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
```

CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 10 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.
XX
SQ Sequence 188 AA;

Query Match 6.3%; Score 443.5; DB 22; Length 188;
Best Local Similarity 46.3%; Pred. No. 1.2e-21;
Matches 89; Conservative 31; Mismatches 66; Indels 5; Gaps 1;

```
Qy     343 LQSWQEQTMGTFLVQQAQRFPGLRDFLFSQLVAQLWRNFDEQQNQRGWALMVILLSSFAF 402
Db       4 LHGTLRMILNGIVNGNWTFFALRDFILALLANQVWFNFNAYNCFFGWILLAACFSGFAF 63

Qy     403 TPALEKPLLKFVSDQAPSGMAALCQHFLLGALEQTPLAPMASRSHPPTQLEWKAGLRRGR 462
Db      64 SPHLCKPLLKFVSDYSQNISFQAVCQHRILQA-MCSGAAPTPPPTQIFWTAIQFKAS 118

Qy     463 MALDVFTPNFRSYSAFVFGWTTGFQFAGWILQFPGILEAPFPGWSVSLHSGDAWPDLPGTD 522
Db     119 MALLVSCFNSFQFSCFVHTWSTGEAVAGDILFHPGLADGWPGWTVAMKNGVQWAFLAGHD 178

Qy     523 FVLDLIGQTF 532
Db     179 YVLDLVSDLE 188
```

RESULT 12
ABB60369
ID ABB60369 standard; Protein; 2167 AA.
XX
AC ABB60369;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7899
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2
XX
PD 27-SEP-2001
XX
PF 23-MAR-2001; 2001WO-US09231
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04472.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 7899; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2167 AA;

Query Match 4.7%; Score 328; DB 22; Length 2167;
Best Local Similarity 21.6%; Pred. No. 1.3e-12;
Matches 162; Conservative 97; Mismatches 263; Indels 228; Gaps 24;

```
Qy       6 QEAEPQAGTEQNKETLASRFQQTLGSLLAPMGSPGHVYYVIHCLNETFSKIPGLLDWGHVA 65
Db     587 QTIEMGAFTPKFTPTLSTQFPYSLDALMPTLSGCQPEFIPIPFNELKFFMMFDFGLCC 647

Qy      66 EQLRQAGILEIIGTRSTHFFVPVSFQVFLAPFHALGSWRQFA ASQLGSQGAILGEVLG 125
Db     648 PQLPYSGMMFTIPIFPAGYPIPHGFPEPVFPYPPLIPQVPPAHPTPQQAATSEIQAVVLG 707

Qy     126 AESPLYHLGVTQVLLQEQGWQQLEQLWAQPRSQALLTLHPGLRACITRQR-IRL------ 176
Db     708 KSD--YQLGHTKVFLKDAHDLFLEQERDFVLTPPILIQPSIPGWVYPPPPLPLRAAAIT 765

Qy     177 ----------------------LPRMQARVR ---  ---  ----GLQARKP-YLQPRSA 199
Db     766 VLPPWPGYAQFFPYFNMPVSYMPLQALIPSFVLSHPFPHLFSHIVRLQAHAFYLVPFFY 825

Qy     200 LGQLNTILLVARPL ---LPPPQKLPGAPGP -  --------HSGPPWGK-VSN 247
Db     826 GHKMWAVIKIQSHVRRMIAMRFYPKLPLPHKQPAPVLQLPKLPEQFLLHPGNVHAPPIAF 885

Qy     248 M -----DLGPLEIPAQLATLLRPAGHQALLT----------GSITESLPPEVPAR 298
Db     886 QHYRDFLHELEPPEIQEQLFN-PPRVEVNMNIINDAAPPQEFFVDPGFLVEAMFDFLPDS 944

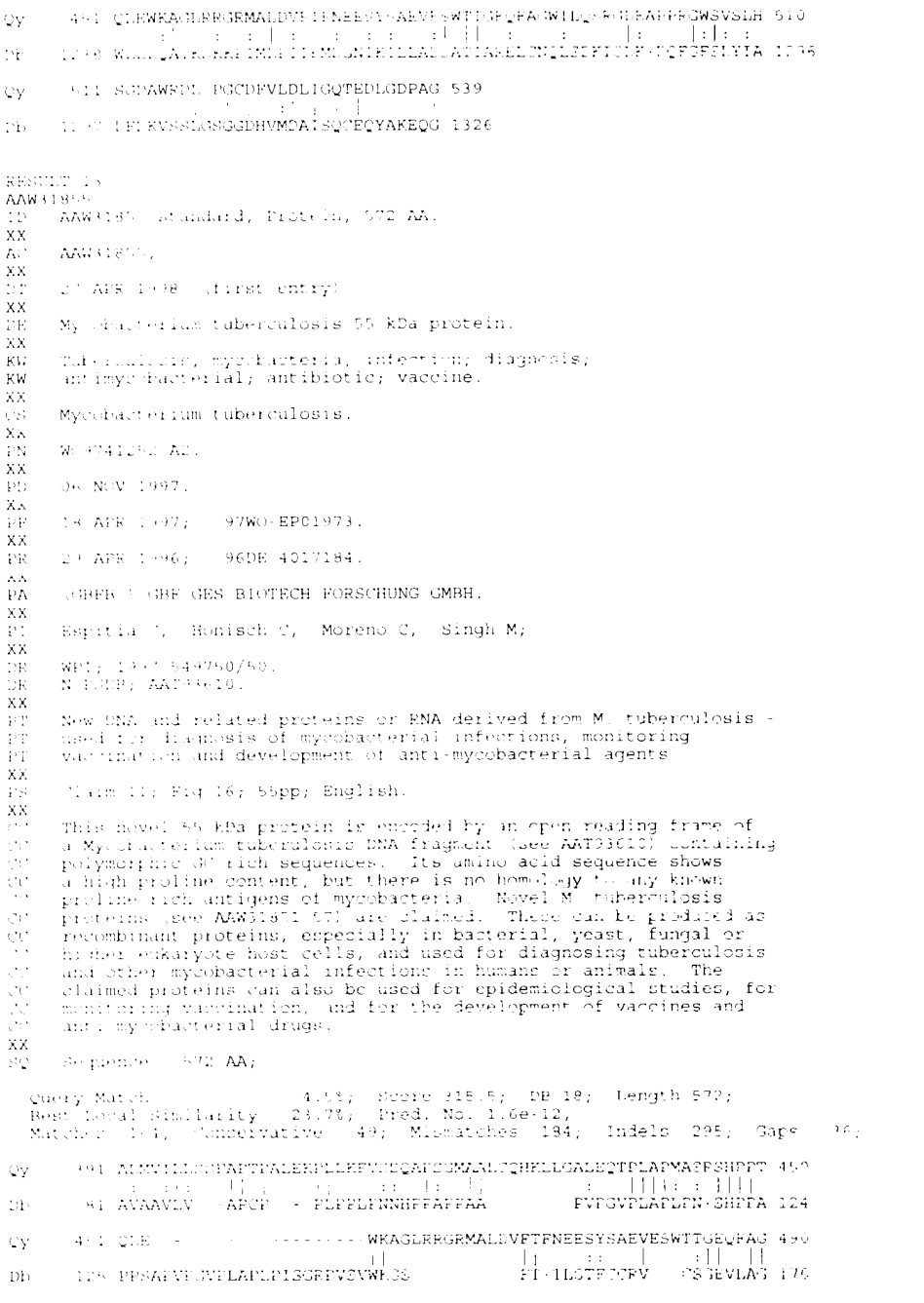
Qy     299 PSLTLPP---------------------------------DIDQFPPSSFVSTSFQ--- 341
Db     945 SSDAPTPHGGRETSVFNDLPHAQNVNQRDIIAPIHISEDEEDLSEFKFQKFAATYFQGNV 1004

Qy     342 ---------KPFLP-------PPGQPLDEPLTRLLGENPQ- ---QALEINRVMLRL 387
Db    1005 NHQYAKKALFHPLLPLHTQSEQLAAQALWTTLRFTGDMFEPKYHTMDRMDTTGVMSKVT 1064

Qy     388 ----------------------------------------------LGE-------- 400
Db    1065 ATLSENFIPSKEPQFAQLMGLMDAFLPQFTFSIFHFLVFLTLIPPEMKLCEDVPRPLQDD 1124

Qy     401 ----GSLQSWQEQTMGT ---PLVQCAQPPPQLPQELFPQLWAQLWRNFDPQQNQRGW 450
Db    1125 EYTADSYQSWLQSKITSNLEFLHFIIGHSILPAPLPEIVTPPLLTNMPLFSSHARGW 1184

Qy     451 ALMVILLSSPAPTPALRKPLLKFVSDQAPSGMAALCQHFLLGALEQTPLAPMASRGHPPT 510
Db    1185 ILLSLCVGCPAPSEKFVNYLPAFIP RSPPGYAPYCEPFLKFTDNN GTRNQPPG 1237
```



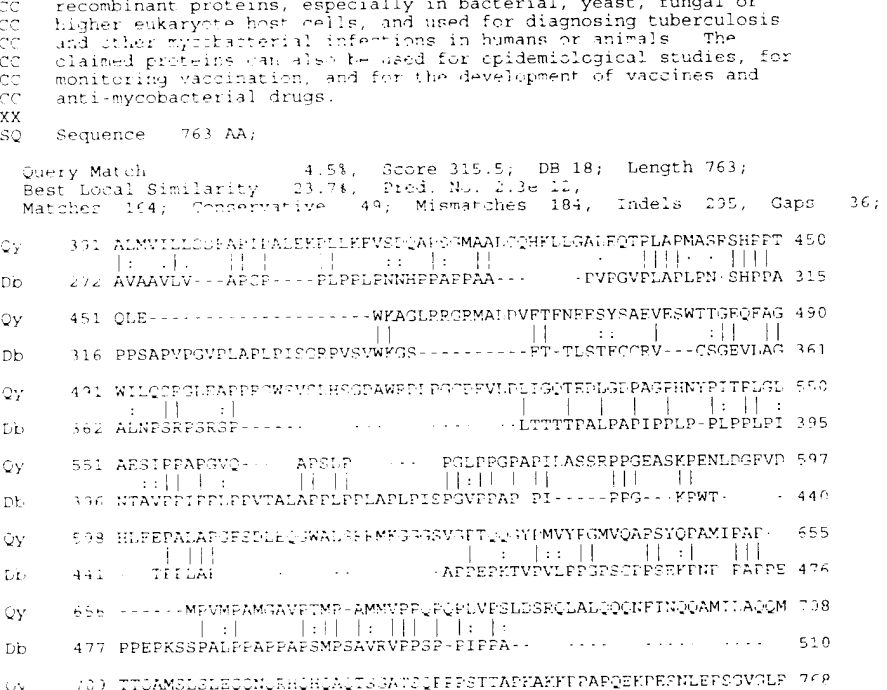
RESULT 13
AAW31855
ID AAW31855 standard; Protein; 572 AA.
XX
AC AAW31855;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 55 kDa protein.
XX
KW Tuberculosis; mycobacteria; infection; diagnosis;
KW antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI Espitia C, Honisch C, Moreno C, Singh M;
XX
DR WPI; 1997-549750/50.
DR N-PSDB; AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT useful in diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
XX
PS Claim 11; Fig 16; 56pp; English.
XX
CC This novel 55 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31871-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
SQ Sequence 572 AA;

Query Match 4.5%; Score 315.5; DB 18; Length 572;
Best Local Similarity 23.7%; Pred. No. 1.6e-12;
Matches 142; Conservative 49; Mismatches 184; Indels 295; Gaps 26;

RESULT 14
AAW31852
ID AAW31852 standard; Protein; 763 AA.
XX
AC AAW31852;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 74 kDa protein.
XX
KW Tuberculosis; mycobacteria; infection; diagnosis;
KW antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI Espitia C, Honisch C, Moreno C, Singh M;
XX
DR WPI; 1997-549750/50.
DR N-PSDB; AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring

PT vaccination and development of anti-mycobacterial agents
XX
PS Claim 5; Fig 13; 55pp; English.
XX
CC This novel 74 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberulosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
SQ Sequence 763 AA;

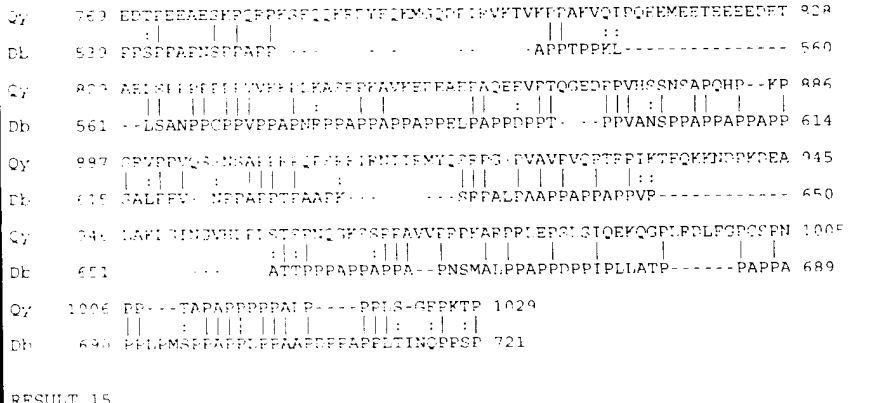
Query Match 4.5%, Score 315.5; DB 18; Length 763;
Best Local Similarity 23.7%, Pred. No. 2.3e-12;
Matches 164; Conservative 49; Mismatches 184; Indels 235; Gaps 36;

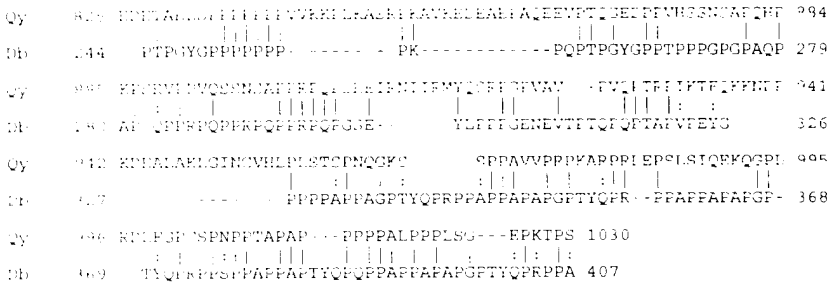


RESULT 15

ABB70063
ID ABB70063 standard; Protein; 446 AA.
XX
AC ABB70063;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 36981.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL14166.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell cell
PT interactions -
XX
PS Disclosure, SEQ ID NO 36981; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABR57737-ABR72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 446 AA;

Query Match 4.4%; Score 310.5; DB 22; Length 446;
Best Local Similarity 24.2%; Pred. No. 2.5e-12;
Matches 126; Conservative 21; Mismatches 161; Indels 213; Gaps 21;





Search completed: January 10, 2003, 02:17:45
Job time : 1.4,419 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: January 1[illegible], [illegible], [illegible] , Search time 47.5262 Seconds
(without alignments)
2704.441 Million cell updates/sec

Title: US-09-803-126-1
Perfect score: 6991
Sequence: 1 M[illegible] WQFKSKLGPAQEVPKTPEVK 1337

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1759	25.2	3511	2	A59295	unconventional myo
2	1710	24.9	1793	2	T42386	unconventional myo
3	1717	24.6	3530	2	A59266	unconventional myo
4	510	7.3	2910	2	T32238	hypothetical prote
5	339	4.8	1189	2	S49915	extensin-like prot
6	316.5	4.5	2659	2	T25898	hypothetical prote
7	312.5	4.5	2175	2	A59255	myosin VIIa, long
8	310	4.4	1611	2	T38236	hypothetical prote
9	303.5	4.3	2215	2	T30870	myosin VIIa - mous
10	[illegible]	4.[illegible]	[illegible]	[illegible]	A[illegible]	myosin VII-like pr
11	290	4.1	1184	2	G01763	atrophin-1 - human
12	289.5	4.1	1184	2	S50842	atrophin-1 human
13	286	4.1	1006	2	T42731	atrophin-1 related
14	[illegible]	4.[illegible]	[illegible]	[illegible]	S[illegible]	hydroxyproline-ric
15	280.5	4.0	3164	1	WMBEH6	UL36 protein - hum
16	276.5	4.0	2142	2	B35098	MHC class III hist
17	276.5	4.0	4957	2	T03455	ALR protein - huma
18	276.5	4.0	5[illegible]	[illegible]	[illegible]	ALR protein - huma
19	276	3.9	1201	2	G86441	unknown protein [i
20	272.5	3.9	617	2	S51342	verprolin - yeast
21	272.5	3.9	1870	2	S37671	MHC class III hist
22	272.5	3.9	197[illegible]	[illegible]	S36102	MHC class III hist
23	269.5	3.9	761	2	C84672	hypothetical prote
24	266.5	3.8	1074	2	T16251	hypothetical prote
25	264	3.8	716	2	T26098	hypothetical prote
26	264	3.8	1267	2	T29265	hypothetical prote
27	263	3.8	2715	2	T13049	eyelid - fruit fly
28	261.5	3.7	2643	2	T51022	hypothetical prote
29	260	3.7	1494	2	T14355	protein-tyrosine-p
30	258.5	3.7	990	2	T14756	hypothetical prote
31	258.5	3.7	7962	2	I38346	elastic titin - hu
32	256.5	3.7	378	2	S14959	proline rich prote
33	256.5	3.7	152[illegible]	2	T00272	hypothetical prote
34	256	3.7	847	2	F96531	hypothetical prote
35	255.5	3.7	727	2	C84534	hypothetical prote
36	252	3.6	1736	2	T00391	hypothetical prote
37	252	3.6	1794	2	T38459	hypothetical diver
38	250.5	3.6	350	2	S22456	hydroxyproline-ric
39	250	3.6	2187	2	T30826	nascent polypeptid
40	249.5	3.6	862	2	T46286	hypothetical prote
41	249.5	3.6	1307	2	T25563	hypothetical prote
42	247.5	3.5	303	2	G28264	hydroxyproline ric
43	247.5	3.5	574	2	T43556	Wiskott Aldrich sy
44	247.5	3.5	574	2	T29915	wiskott-aldrich sy
45	247.5	3.5	760	2	T06291	extensin homolog T

ALIGNMENTS

RESULT 1
A59295
unconventional myosin-15 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59295
R;Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; Mil
an, T.B.; Fridell, R.A.
Genomics 61, 243-258, 1999
A;Title: Characterization of the human and mouse unconventional myosin XV genes responsi
A;Reference number: A59266; MUID:20021702; PMID:10552926
A;Accession: A59295
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3511 <LIA>
A;Cross-references: GB:AF144035; NID:g6714084; PIDN:AAF[illegible].1; PID:g[illegible]
C;Genetics:
A;Gene: MGI:Myo15
A;Cross-reference: MGI:1261811
A;Map position: 11:33.9
C;Superfamily: myosin motor domain homology
F;1209-1871/Domain: myosin motor domain homology <MMO>

Query Match 25.2%; Score 1759; DB 2; Length 3511;
Best Local Similarity 32.1%; Pred. No. 9.3e-75;
Matches 482; Conservative 225; Mismatches 402; Indels 308; Gaps 32;

Qy 6 QEAEPQAGTEQN-- ---KPTLAGPPQLTLGPLLAPLGSPGHVYVTHCLNPTPGKIPGLL 59
Db 1733 QTAPPRLGKSSSIIRLYKAHTVAARPQQSLLDLVPPQ PPCHPLPVPCLVPPHVPPPGLP 1791

Qy 60 DVGHVAEQLRQAGILEIIGTPSTHPEVPVSPQVPLAPPHALGGPPIPAAGPQPPGGAILG 119
Db 1792 EPDVMMAQLKYSGVLETVRIPPEJPPVPIPPQVPIPPVP LVALPLDVPAPPPMGVGLLP 1851

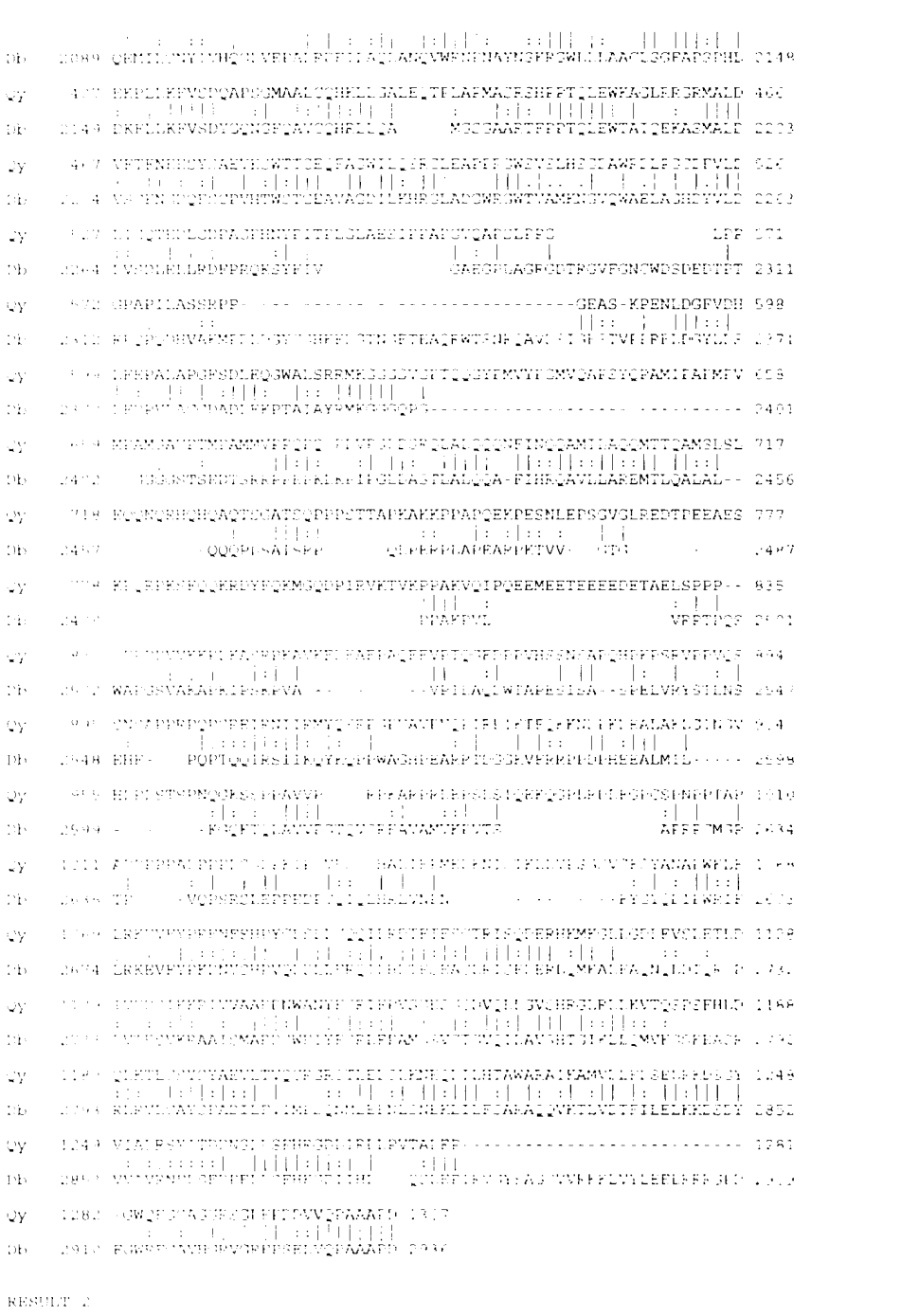
Qy 120 EVLGAESPLYHLGVTQVLLQEQGWQQLGSLWAQPRSQALLTLHRGLRACITPQPLRLLPR 179
Db 1852 ELGPVTPLMYPVPIGPPLPPPHLPLLPLMPPEVQPPAALTPPPLPPPIGPHPPGLRR 1911

Qy 180 ----MQARVRGLQAPKRYLQRRGALGQLNTLL-LVAPPLLPPPQPIPGAPGPHSGPPW- 232
Db 1912 PIILLQSPAPGPLAPQPYGQMPQGALGPPPSLVHTYVP -PPYLMTPAPGPPPAQPPWI 1969

Qy 233 ------GKVSNMDLGRLEIPAQLAILLERAEGHQALLTGSITESLPPEVPARPGLTLPPP 286
Db 1969 PSQPPLSPPPEVVPVPHLEVIASVA[illegible]PPVVTPAPPLQAPPGVTLPLD 2028

Qy 287 IDQPTPGSPVSTSPQVPPLPPIPQPLPPMTPLPSENIQQALEINRVKLELLGEGGLQSW 346
Db 2029 INNYTMAAPIPGHPPPPPPGMLTVPLPMPLTPLPVPHHAPAIGVPKLILLRFMGDPHLHGT 2088

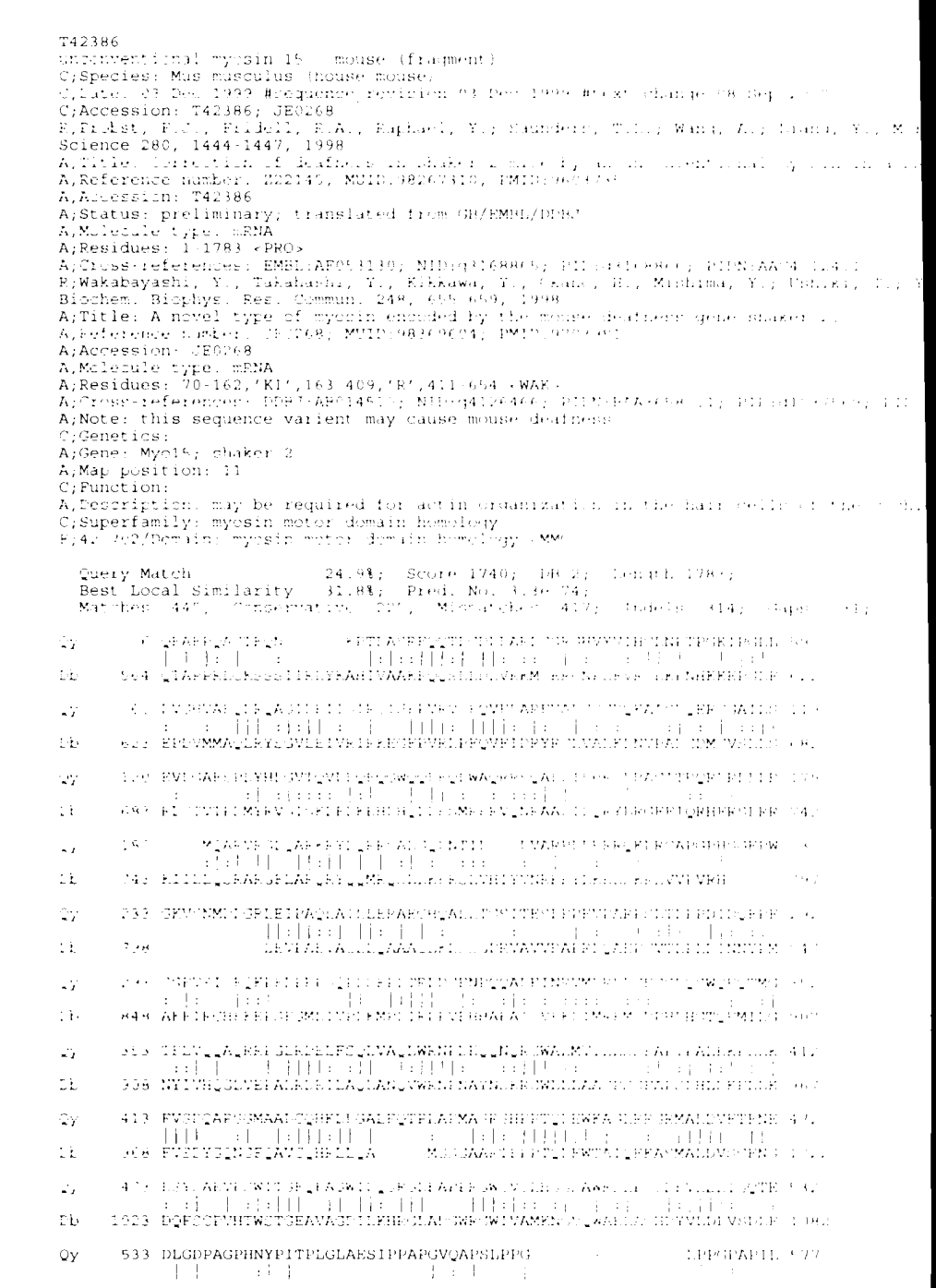
Qy 347 QEQTMGTPLVQQAQRPGLRDELPSQLVAQLNRNTDEQQVQPGWALMVLLSSPAPTPAL 406

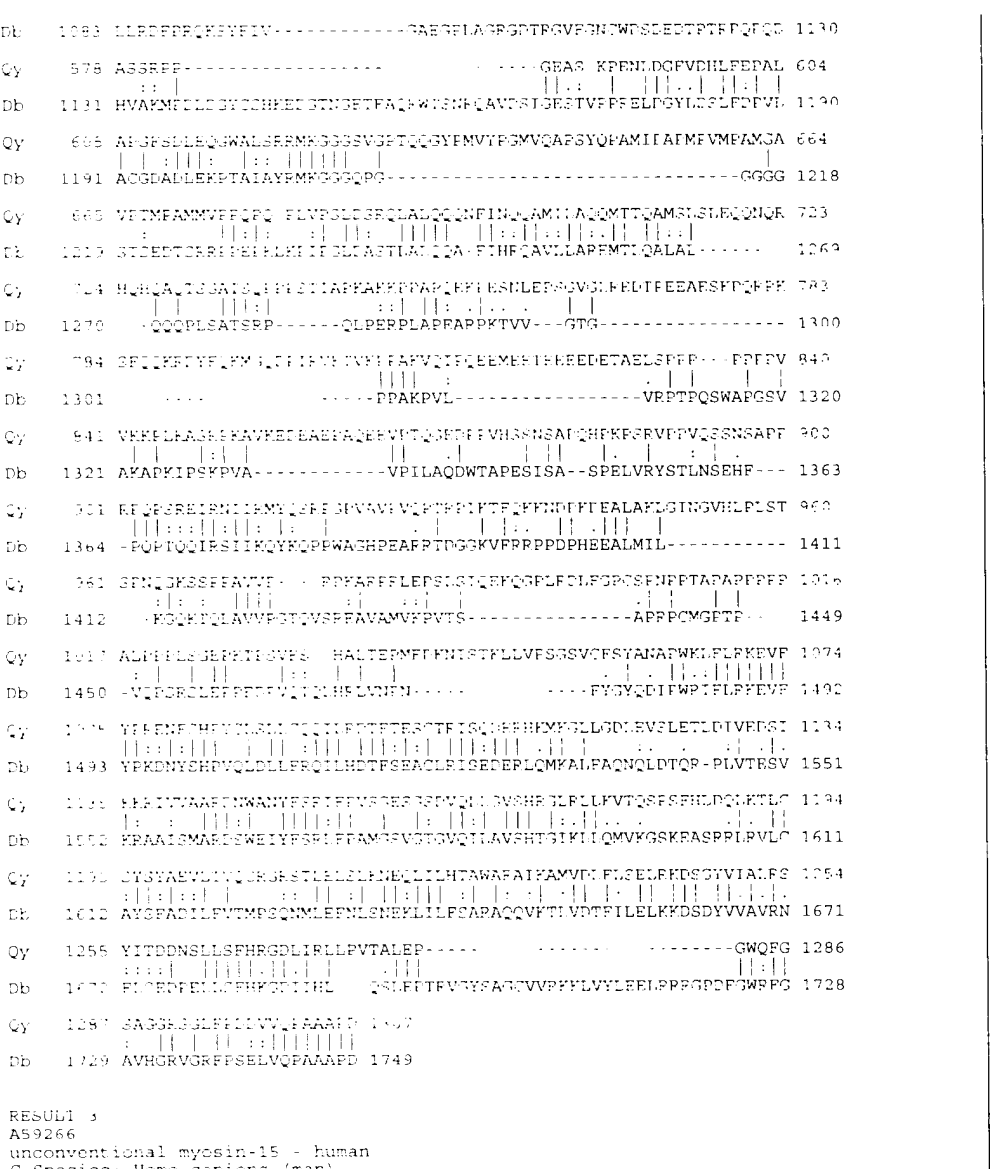


RESULT 2

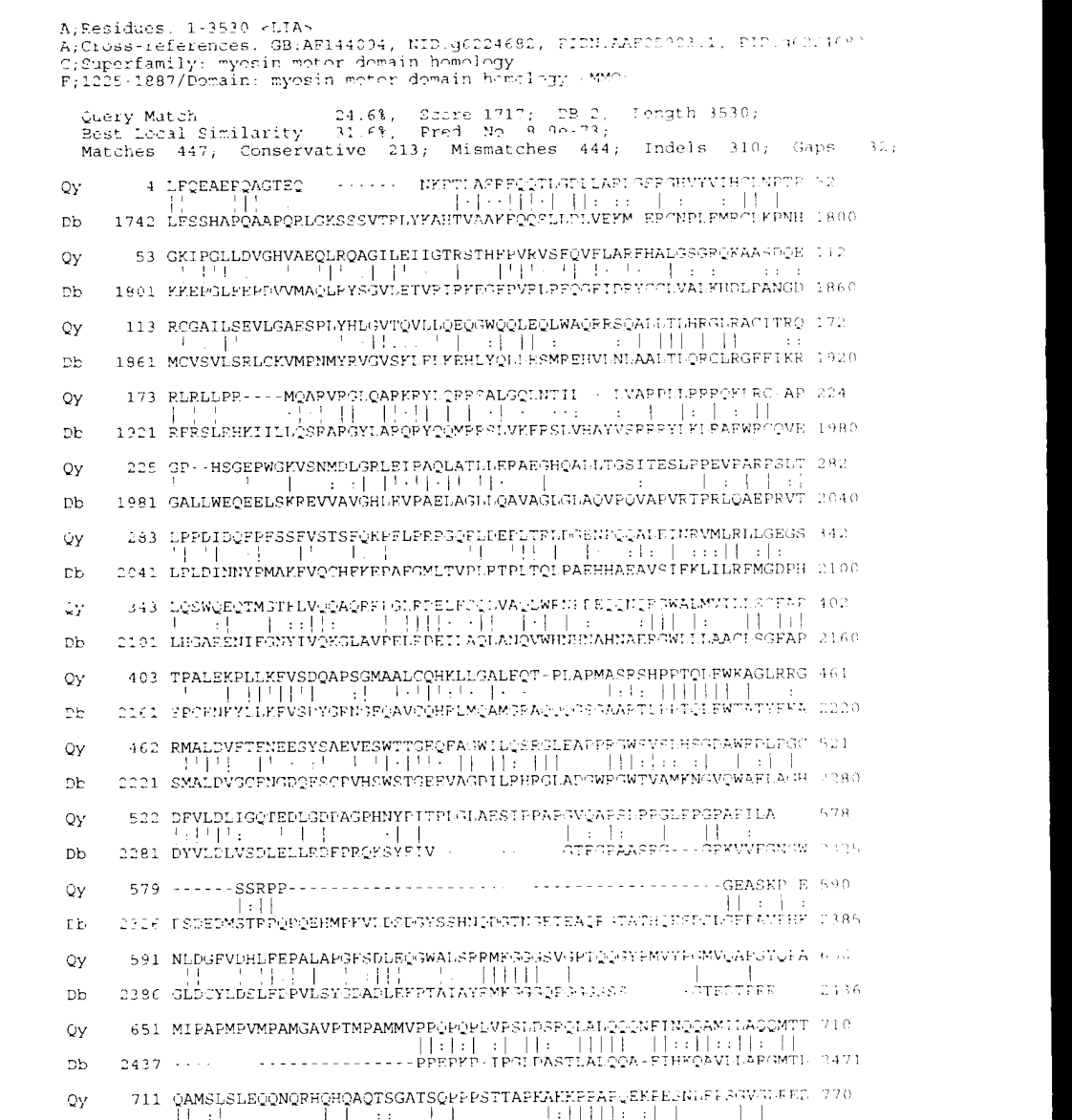
T42386
unconventional myosin 15 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 23-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
C;Accession: T42386; JE0268
R;Fidiat, F.J., Fridell, R.A., Rupert, V.; Sheppard, T.L.; Ward, A.; Irani, V., M.
Science 280, 1444-1447, 1998
A;Title: Correction of deafness in shaker 2 mice by an unconventional myosin in a novel family of mammalian motors
A;Reference number: A02145; MUID:98267819; PMID:9603743
A;Accession: T42386
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1783 <PRO>
A;Cross-references: EMBL:AF053140; NID:g2688056; PID:g2688057; PIDN:AAC53
R;Wakabayashi, Y., Takahashi, Y., Kikkawa, Y., Kano, H., Mishima, Y.; Ushiki, T.;
Biochem. Biophys. Res. Commun. 248, 655-659, 1998
A;Title: A novel type of myosin encoded by the mouse deafness gene shaker-2
A;Reference number: JE0268; MUID:98269604; PMID:9703
A;Accession: JE0268
A;Molecule type: mRNA
A;Residues: 70-162,'KI',163-409,'R',411-654 <WAK>
A;Cross-references: DDBJ:AB014511; NID:g4126466; PID:BAA03
A;Note: this sequence variant may cause mouse deafness
C;Genetics:
A;Gene: Myo15; shaker 2
A;Map position: 11
C;Function:
A;Description: may be required for actin organization in the hair cells of the
C;Superfamily: myosin motor domain homology
F;4-302/Domain: myosin motor domain homology <MM>

Query Match 24.9%; Score 1740; DB 2; Length 1783;
Best Local Similarity 31.8%; Pred. No. 3.9e-94;
Matches 446; Conservative 227; Mismatches 417; Indels 414; Gaps 41;



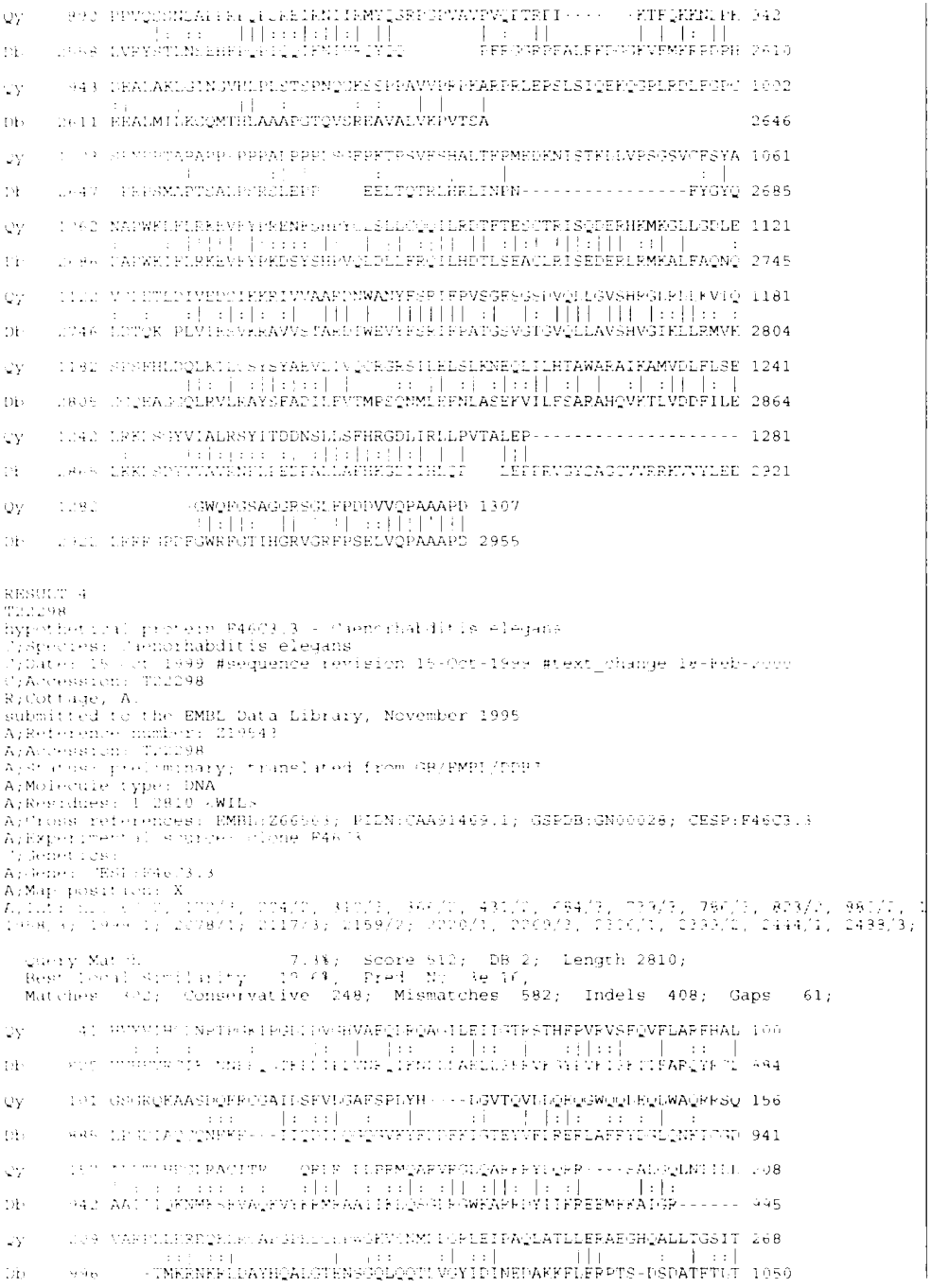


RESULT 3
A59266
unconventional myosin-15 - human
C;Species: Homo sapiens (man)
C;Date: [illegible] #sequence_revision [illegible] #text_change 08-Sep-2000
C;Accession: A59266
R;Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W., Probst, F.J., Barber, T.D., Mi
an, T.B.; Friedell, R.A.
Genomics 61, 243-258, 1999
A;Title: Characterization of the human and mouse unconventional myosin XV genes responsi
A;Reference number: A59266; MUID:20001762; PMID:10552926
A;Accession: A59266
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA



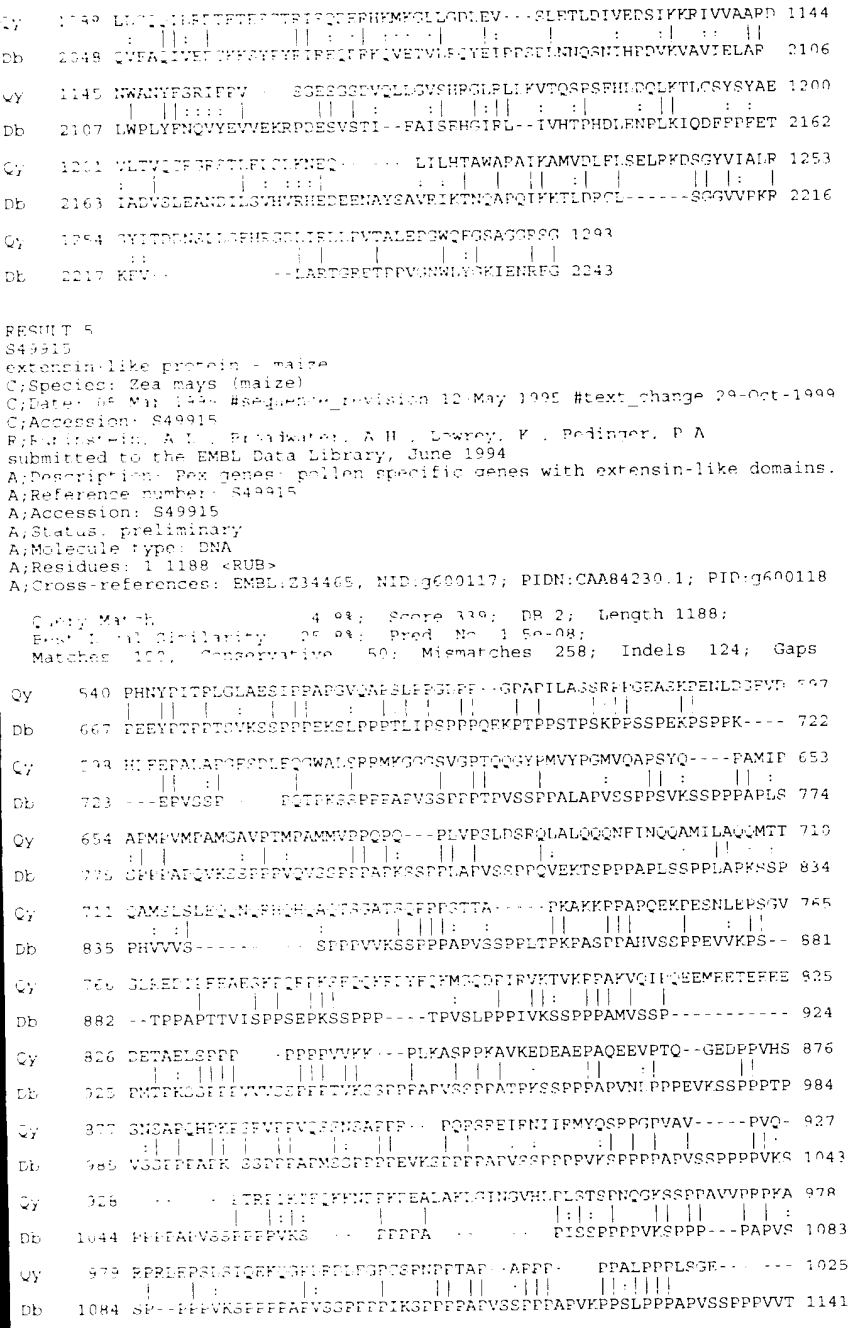
A;Residues: 1-3530 <LIA>
A;Cross-references: GB:AF144004; NID:g6224682; PIDN:AAF05203.1; PID:g6224683
C;Superfamily: myosin motor domain homology
F;1225-1887/Domain: myosin motor domain homology <MMO>

Query Match 24.6%; Score 1717; DB 2; Length 3530;
Best Local Similarity 31.6%; Pred. No. 8.9e-73;
Matches 447; Conservative 213; Mismatches 444; Indels 310; Gaps 32;



RESULT 4
T22298
hypothetical protein F46C3.3 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T22298
R;Cottage, A.
submitted to the EMBL Data Library, November 1999
A;Reference number: Z20943
A;Accession: T22298
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2810 <WIL>
A;Cross-references: EMBL:Z66035; PIDN:CAA91409.1; GSPDB:GN00028; CESP:F46C3.3
A;Experimental source: clone F46C3
;Gene: "ESI:F46C3.3
A;Map position: X
A;[1]: 1-2, 173/3, 224/3, 310/3, 367/3, 431/3, 484/3, 729/3, 780/3, 803/2, 981/3, 1
1408/3; 1444/1; 2078/1; 2117/3; 2159/2; 2200/1, 2269/3, 2316/1, 2353/2, 2444/1, 2488/3;

Query Match 7.4%; Score 512; DB 2; Length 2810;
Best Local Similarity 19.6%; Pred. No. 3e-16;
Matches 492; Conservative 248; Mismatches 582; Indels 408; Gaps 61;

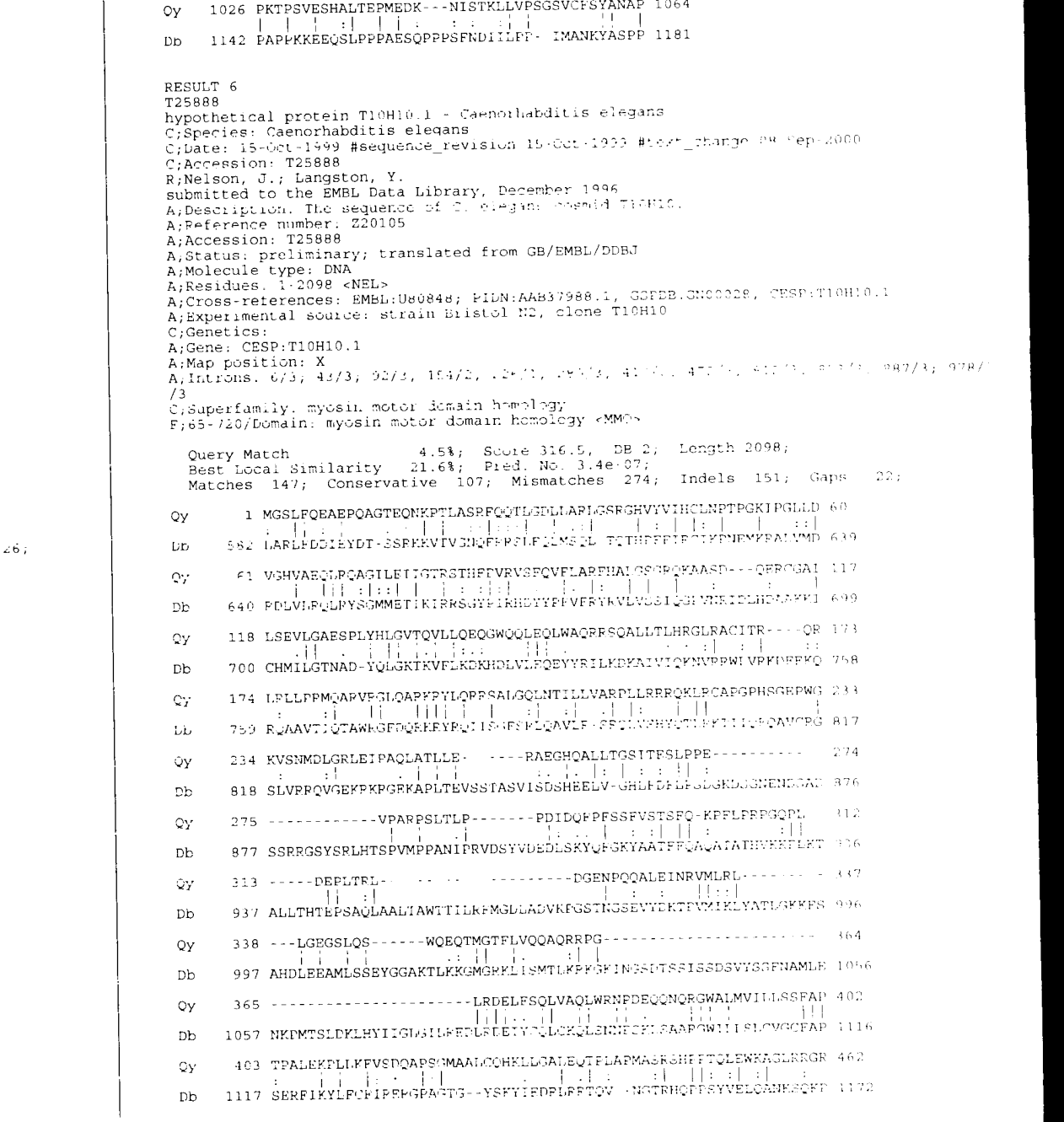


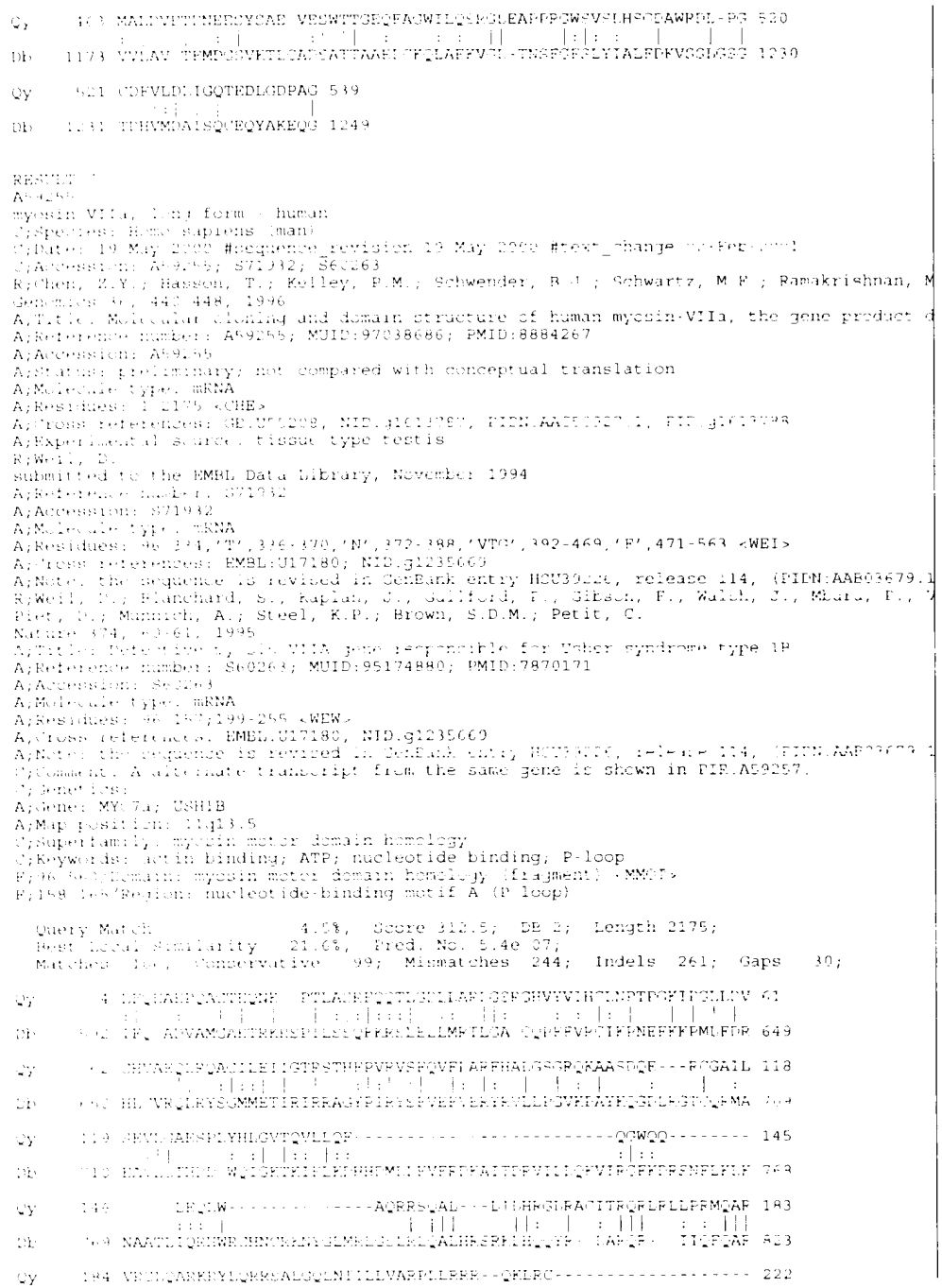
RESULT 5
S49915
extensin-like protein - maize
C;Species: Zea mays (maize)
C;Date: 05-May-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C;Accession: S49915
R;Kieliszewski, M.J., Brandwein, A.H., Lewroy, K., Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A;Description: Pex genes: pollen specific genes with extensin-like domains.
A;Reference number: S49915
A;Accession: S49915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: EMBL:Z34465, NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 4.9%; Score 339; DB 2; Length 1188;
Best Local Similarity 25.8%; Pred. No. 1.5e-08;
Matches 192; Conservative 50; Mismatches 258; Indels 124; Gaps 26;

RESULT 6
T25888
hypothetical protein T10H10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T25888
R;Nelson, J.; Langston, Y.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid T10H10.
A;Reference number: Z20105
A;Accession: T25888
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2098 <NEL>
A;Cross-references: EMBL:U80848; PIDN:AAB37988.1; GSPDB:GN00028, CESP:T10H10.1
A;Experimental source: strain Bristol N2, clone T10H10
C;Genetics:
A;Gene: CESP:T10H10.1
A;Map position: X
A;Introns: 6/3; 43/3; 92/3; 154/2; 254/1; 295/3; 417/1; 472/1; 511/1; 613/1; 987/1; 978/2; 1026/3
C;Superfamily: myosin motor domain homology
F;65-720/Domain: myosin motor domain homology <MMO>

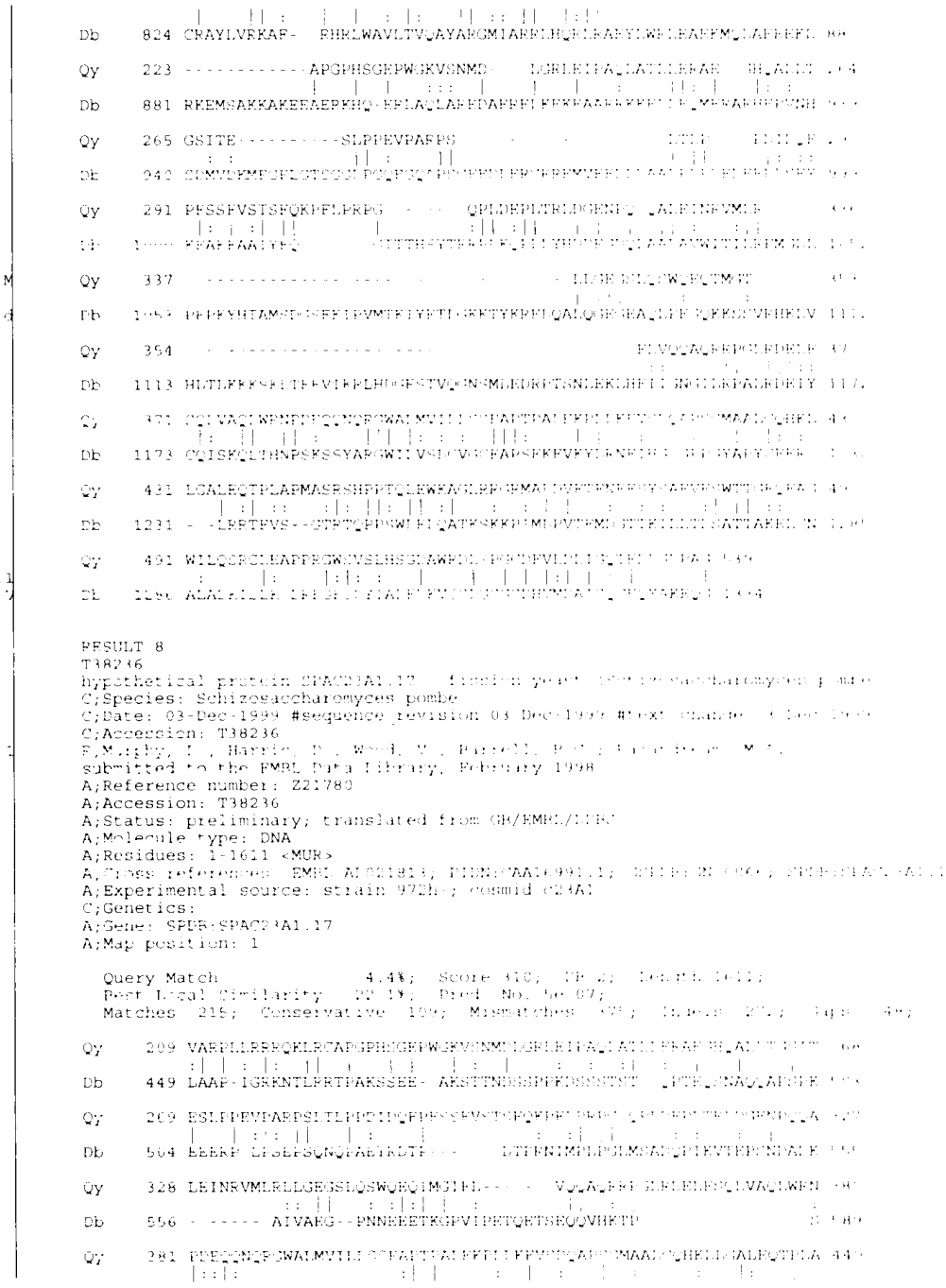
Query Match 4.5%; Score 316.5; DB 2; Length 2098;
Best Local Similarity 21.6%; Pred. No. 3.4e-07;
Matches 147; Conservative 107; Mismatches 274; Indels 151; Gaps 22;





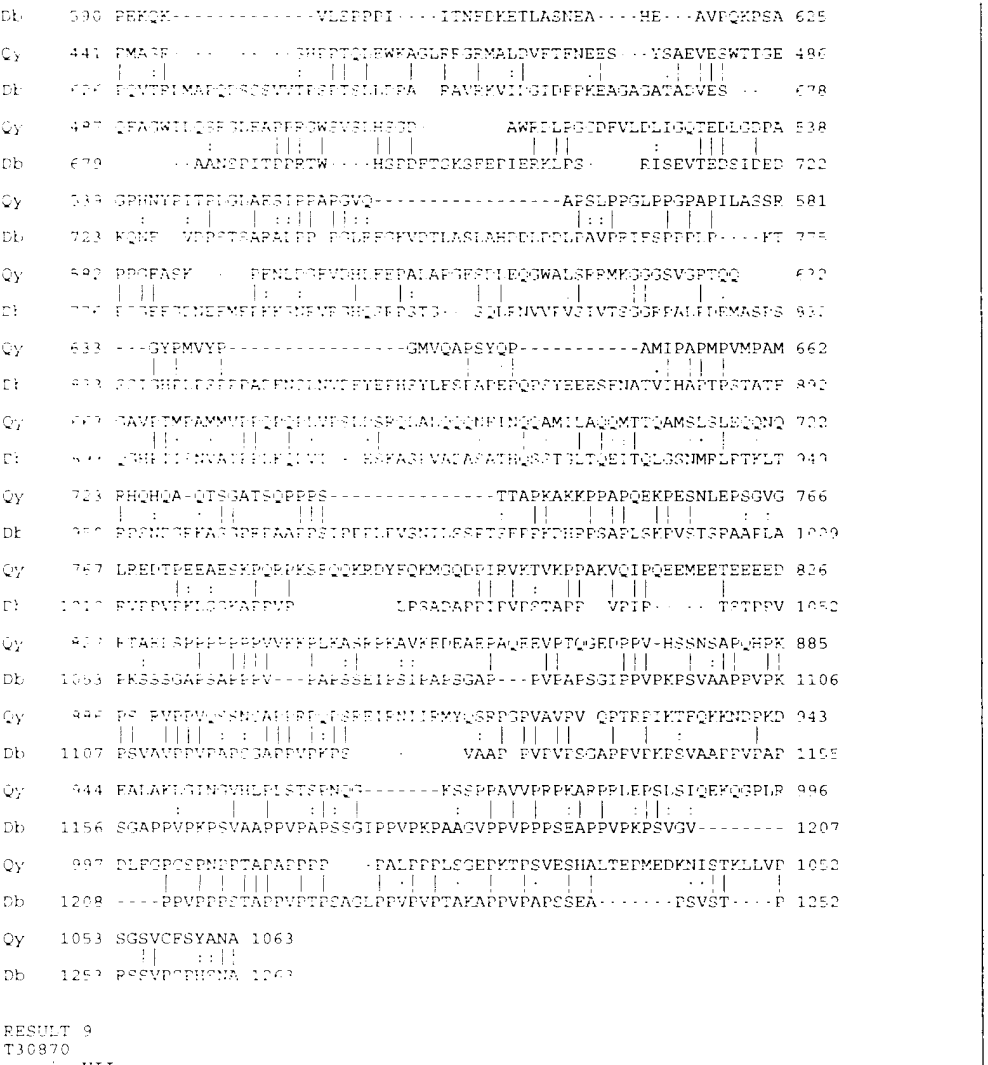
RESULT 7
A59256
myosin VIIa, long form - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jun-2000
C;Accession: A59256; S71232; S60263
R;Chen, Z.Y.; Hasson, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M
Genomics 34, 341-348, 1996
A;Title: Molecular cloning and domain structure of human myosin-VIIa, the gene product
A;Reference number: A59256; MUID:97038686; PMID:8884267
A;Accession: A59256
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-2175 <CHE>
A;Cross-references: GB:U17208; NID:g1617787; PIDN:AAB03679.1; PID:g1617788
A;Experimental source: tissue type testis
R;Weil, D.
submitted to the EMBL Data Library, November 1994
A;Reference number: S71232
A;Accession: S71232
A;Molecule type: mRNA
A;Residues: 96-134,'T',136-170,'N',172-388,'VT',392-469,'P',471-563 <WEI>
A;Cross-references: EMBL:U17180; NID:g1235669
A;Note: the sequence is revised in GenBank entry HCU39226, release 114, (PIDN:AAB03679.1
R;Weil, D.; Blanchard, S.; Kaplan, J.; Guilford, P.; Gibson, F.; Walsh, J.; Mburu, P.; P
Petit, C.; Mannich, A.; Steel, K.P.; Brown, S.D.M.; Petit, C.
Nature 374, 60-61, 1995
A;Title: Defective myosin VIIA gene responsible for Usher syndrome type 1B
A;Reference number: S60263; MUID:95174890; PMID:7870171
A;Accession: S60263
A;Molecule type: mRNA
A;Residues: 96-167;199-265 <WEW>
A;Cross-references: EMBL:U17180; NID:g1235669
A;Note: the sequence is revised in GenBank entry HCU39226, release 114, (PIDN:AAB03679.1
C;Comment: A alternate transcript from the same gene is shown in PIR:A59257.
C;Genetics:
A;Gene: MY07A; USH1B
A;Map position: 11q13.5
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; ATP; nucleotide binding; P-loop
F;96-503/Domain: myosin motor domain homology (fragment) <MMOT>
F;158-165/Region: nucleotide-binding motif A (P loop)

  Query Match 4.5%; Score 312.5; DB 2; Length 2175;
  Best Local Similarity 21.6%; Pred. No. 5.4e-07;
  Matches 104; Conservative 99; Mismatches 244; Indels 261; Gaps 30;



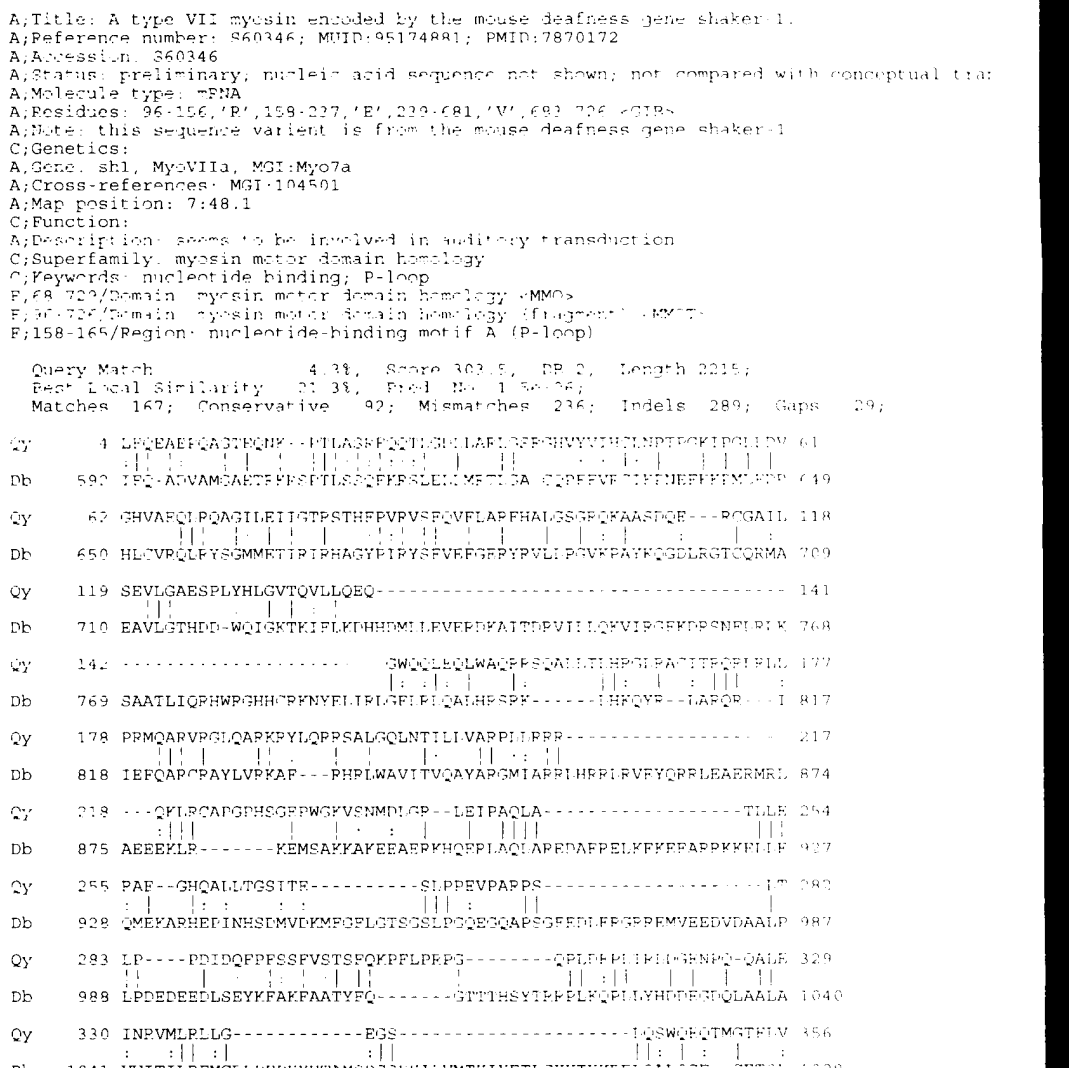
RESULT 8
T38236
hypothetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 14-Dec-2000
C;Accession: T38236
R;Murphy, L.; Harris, D.; Wood, V.; Rutherford, K.; Barrell, B.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21780
A;Accession: T38236
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1611 <MUR>
A;Cross-references: EMBL:AL021813; PIDN:CAA19991.1; GeneID:2539966; SPDP:SPAC23A1.17
A;Experimental source: strain 972h-; cosmid c23A1
C;Genetics:
A;Gene: SPDB:SPAC23A1.17
A;Map position: 1

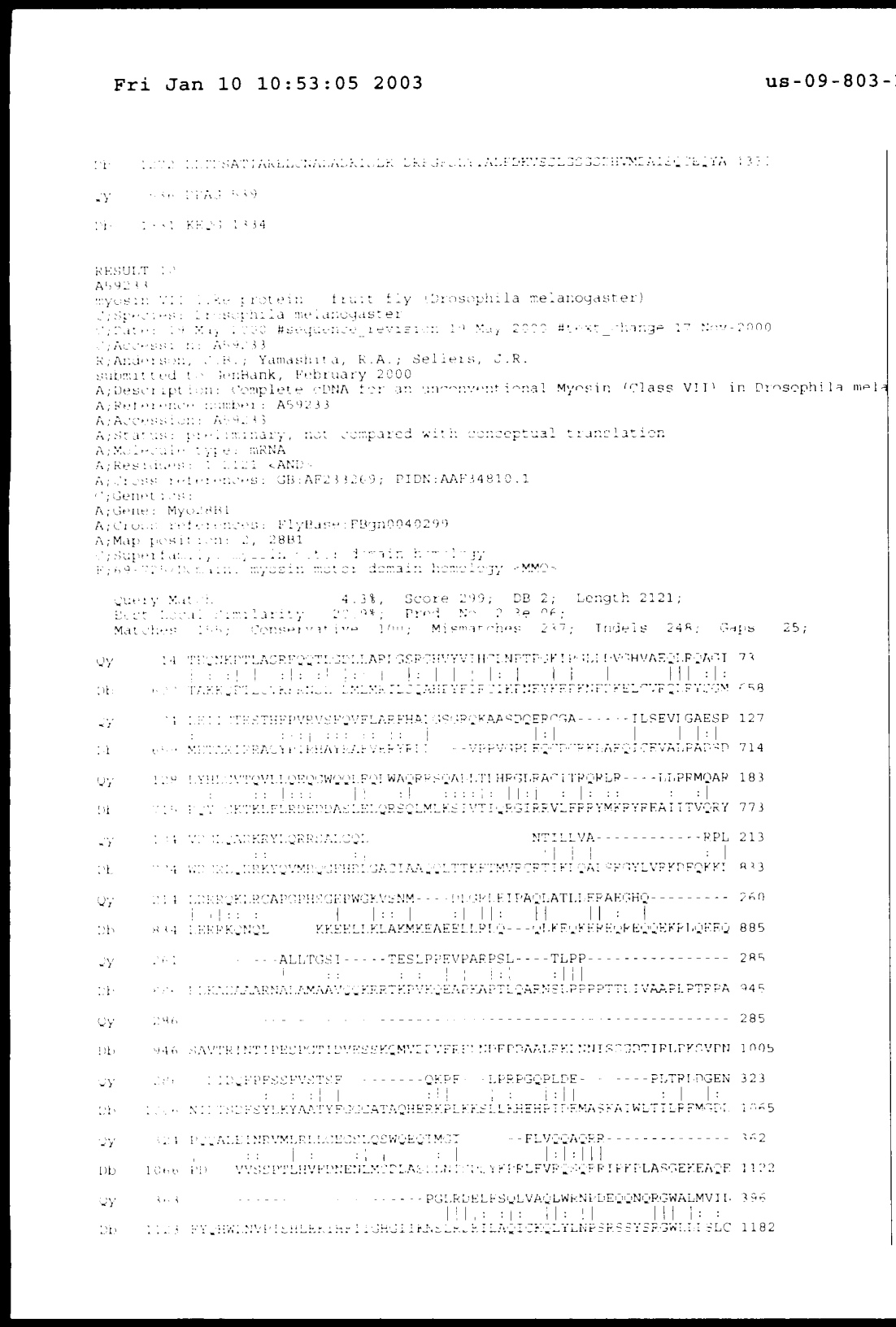
  Query Match 4.4%; Score 310; DB 2; Length 1611;
  Best Local Similarity 22.1%; Pred. No. 5e-07;
  Matches 215; Conservative 109; Mismatches 423; Indels 276; Gaps 46;



RESULT 9
T30870
myosin VIIa - mouse
N;Alternate names: gene shaker 1 protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Feb-2001
C;Accession: T30871, S60346
R;Walsh, J ; Mburu, P ; Gibson, F ; Liu, X ; Saw, D.; Steel, K P ; Brown, S D.M.
submitted to the EMBL Data Library, December 1994
A;Description: Mutation analysis of the mouse myosin VIIa deafness gene
A;Reference number: S60343
A;Accession: T30870
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2215 <WAL>
A;Cross-references: EMBL:U81462; NID:g1778380; PID:g1778382; PIDN:AAB40708.1
R;Gibson, F ; Walsh, J ; Mburu, P ; Varela, A ; Brown, K.A.; Antonio, M.; Beisel, K.W ; Steel, K.P.; Brown, S.D.M.
Nature 374, 62-64, 1995

A;Title: A type VII myosin encoded by the mouse deafness gene shaker-1.
A;Reference number: S60346; MUID:95174881; PMID:7870172
A;Accession: S60346
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 96-156,'R',158-227,'E',229-681,'V',683-726 <GIB>
A;Note: this sequence variant is from the mouse deafness gene shaker-1
C;Genetics:
A;Gene: sh1; MyoVIIa; MGI:Myo7a
A;Cross-references: MGI:104501
A;Map position: 7:48.1
C;Function:
A;Description: seems to be involved in auditory transduction
C;Superfamily: myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;68-720/Domain: myosin motor domain homology <MMO>
F;50-726/Domain: myosin motor domain homology (fragment) <MYT>
F;158-165/Region: nucleotide-binding motif A (P-loop)





RESULT 10
A59233
myosin VII [imported] - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 10-May-2000 #sequence_revision 10-May-2000 #text_change 17-Nov-2000
C;Accession: A59233
R;Anderson, C.B.; Yamashita, R.A.; Sellers, J.R.
submitted to GenBank, February 2000
A;Description: Complete cDNA for an unconventional Myosin (Class VII) in Drosophila mela
A;Reference number: A59233
A;Accession: A59233
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-2121 <AND>
A;Cross-references: GB:AF233269; PIDN:AAF34810.1
C;Genetics:
A;Gene: MyoVIIB1
A;Cross-references: FlyBase:FBgn0040299
A;Map position: 2, 28B1
C;Superfamily: myosin motor domain homology
F;8-2121/Domain: myosin motor domain homology <MMO>

Query Match 4.1%, Score 299; DB 2; Length 2121;
Best Local Similarity 22.0%; Pred. No. 2.3e-06;
Matches 156; Conservative 100; Mismatches 237; Indels 248; Gaps 25;

Qy 14 TFQNKPTLACRFQQTLQSLLAPLGSPCHVYVFHTLNPTPSKIPCLIVGHVAEQLPQAGI 73
Db 627 FAPPLPTLGPAPAGLA LMLMATLQAHPVFIPCIRTNPYFPPDPTPELCVPQLFYQGM 686

Qy 74 LEILTESTHPPVPVSFQVFLAPPHALGSGRQKAASDQEPCGA------ILSEVIGAESP 127
Db 687 MSTGKIPRALYPLHATLAAPVAPYPI --VPPVGPLFQVTCPRLAPQICFVALPAPSD 744

Qy 128 LYHLGVTQVLLQEQQWVQLPQIWAQPPSQALLTLHPGLRACTTPQPLR----LLPPMQAP 183
Db 745 PGY DETKLFLRDEPDASLELQRSQLMLKSIVTLPGIREVLFPPYMPPYPEAIITVQRY 773

Qy 184 VGLPPPKRVLQRPGALCQL NTILLVA------RPL 213
Db 774 WDSLPPKVQVMPLQPHPLGACIAALQLTTPFTMVFCPTIFLQALSPGYLVPKDPQKKT 833

Qy 214 LDERQKLRCAPGPHSGPPWGKVSNM----PLGPLEIPAQLATLIPPAEGHQ--------- 260
Db 834 LPPPKQNQL KKPPLLKLAKMKEAEPLLPLQ---QLKPLKPPEQMPQQPKPLQPPQ 885

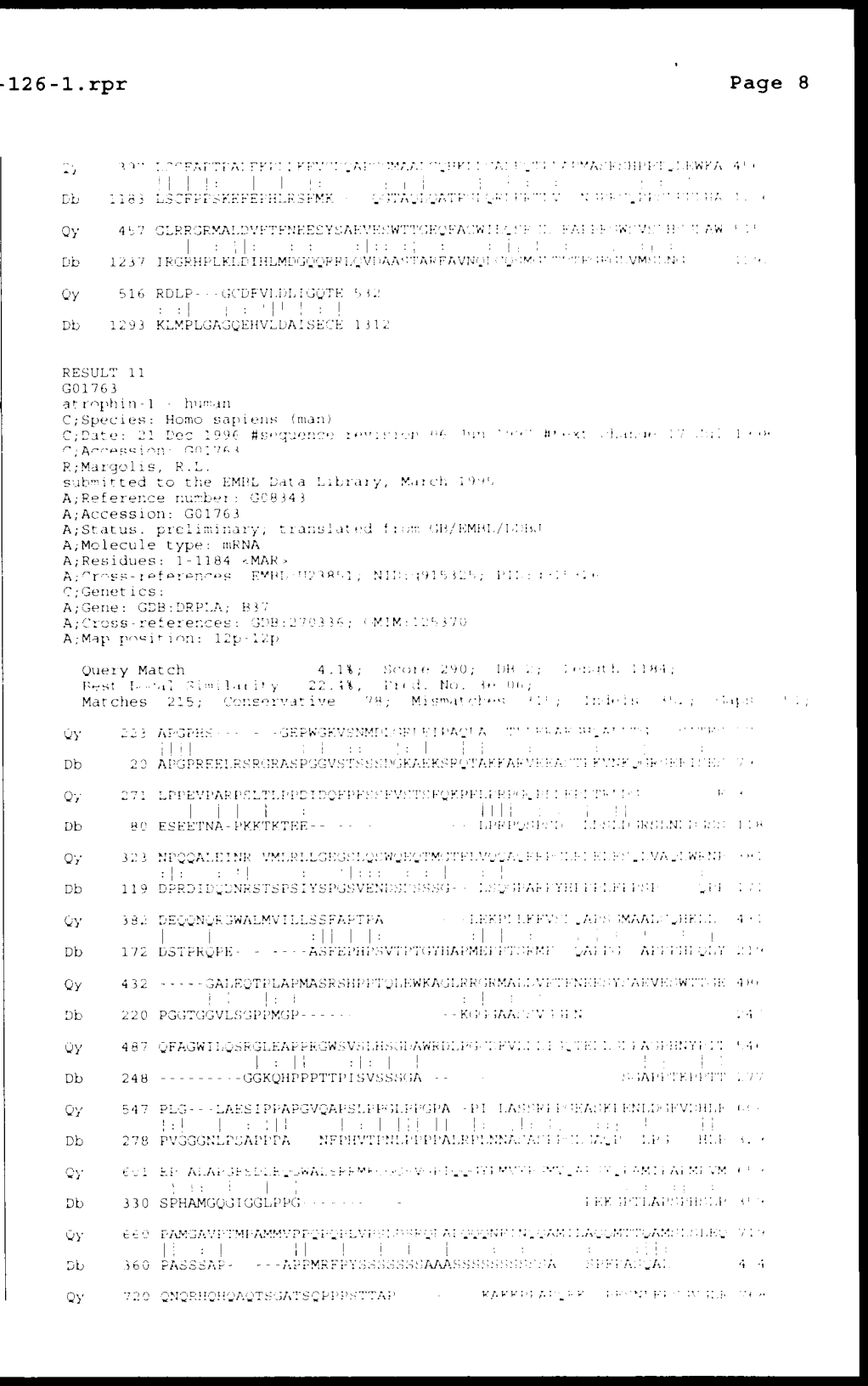
Qy 261 ---ALLTGSI-----TESLPPEVPARPSL----TLPP---------- 285
Db 886 LLKNDAIARNALAMAAVCQPRRTKPVKQEAPPAPTLQARNSLPPPPTTLIVAAPLPTPPA 945

Qy 286 ---------- 285
Db 946 SAVTRINTIPEQPQTIDVPSSKQMVSIVPPIMPPPDAALPKLNNISQQDTIPLPKSVPN 1005

Qy 286 TIDQPPPSGPVSTSP -------QKPP -LPPPGQPLDE- - ----PLTPIDGEN 323
Db 1006 NIITDPSYLRYAATYPPQCATAQHERKPLPPSLLRHEHPIPPMASKAIWLTILPPMPDL 1065

Qy 324 PVQALEINPVMLRLLGDGSLQSWQEQTMGT --PLVQQAQRP---------- 362
Db 1066 PD VVSGPTLHVPDNEDLMDTLAALLNPTGLYPIPLPVPQPPRIPPPLASGEPEAQP 1122

Qy 363 ------ ------PGLRDELPSQLVAQLWKNPDEQNQRGWALMVII 396
Db 1123 PYQHWLNVPISHLEKIMPIIGHGIPANSLRLPILAQICRGLYLNPSPSSYSPGWLIISLC 1182



RESULT 11
G01763
atrophin-1 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-2000 #text_change 17-Jul-2000
C;Accession: G01763
R;Margolis, R.L.
submitted to the EMBL Data Library, March 1996
A;Reference number: G08343
A;Accession: G01763
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1184 <MAR>
A;Cross-references: EMBL:U23851; NID:g915825; PID:g915826
C;Genetics:
A;Gene: GDB:DRPLA; B37
A;Cross-references: GDB:270336; OMIM:125370
A;Map position: 12p-12p

Query Match 4.1%; Score 290; DB 2; Length 1184;
Best Local Similarity 22.4%; Pred. No. 4e-06;
Matches 215; Conservative 98; Mismatches 414; Indels 45; Gaps 11;

Qy 333 APGPHS--- - -GEPWGFVSNMDLGPLEIPAQLA TLIPPAEGHQAPAPPS ----PPPP---
Db 20 APGPREELRSRGRASPGGVSTSSSINGKAEKSPQTAPPAPVPPAQTLPVTPPAPQAPPLPEV 77

Qy 271 LPPEVPARPSLTLPPDIDQPPPSGPVSTSPQKPPLLPPGQLPLPPLPPLPPP 322
Db 80 ESEETNA-PKKTKTEE-- - - - - LPPPGSPGD LDSLDRGLNELPSS 130

Qy 323 NPQQALEINR VMLRLLGDGSLQSWQEQTMGTPLVQQAQRPLPLPLPLPLPPLPVAGLWKNP 382
Db 119 DPRDIDQLNRSTSPSIYSPGSVENSSTSSSG-- LSQGHAPTYHPTPTLPTPP LPT 172

Qy 383 DEQQNQRGWALMVILLSSFAPTPA - LEKPLLPFVSLAPSGMAALQLPELL 432
Db 172 DSTPRQPE- - ----ASPEPHPSVTPTGYHAPMEPTSPMT QAIPG APTPHQAY 219

Qy 432 -----GALEQTPLAPMASRSHPPTQLEWKAGLRPGRMALLVPTPNEEYLAPVEGWTTGE 486
Db 220 PGGTGGVLSGPPMGP------ --KGSGAAGVVHN 254

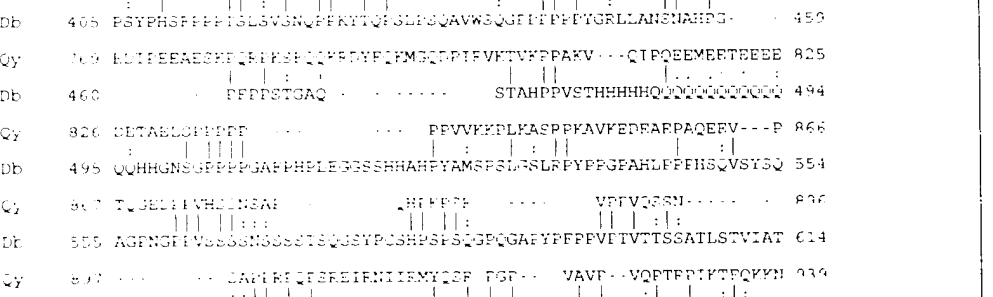
Qy 487 QFAGWILQSRGLEAPPRGWSVSLRSGDAWRDLPGSTPVLIISLTELLTLAGHHNTFIT 546
Db 248 ---------GGKQHPPPTTPISVSSSGA -- GAPITPPTT 272

Qy 547 PLG---LAESIPPAPGVQAPSLPPGLPPGPA -PI LASSPIPEAPKIENLDPFVSHLP 603
Db 278 PVGGGNLPGAPPDA NFPHVTPNLPTPPALRPLNNAGAGIPSLAQP LPT HGL 334

Qy 604 EP ALAPGPSDLPLGWALSPPMPGSVVHLGGTPMVSPVQLASVPLPAMGLAGMPVM 660
Db 330 SPHAMGQGIGGLPPG------ - IPMGTLAPSPHPGLP 369

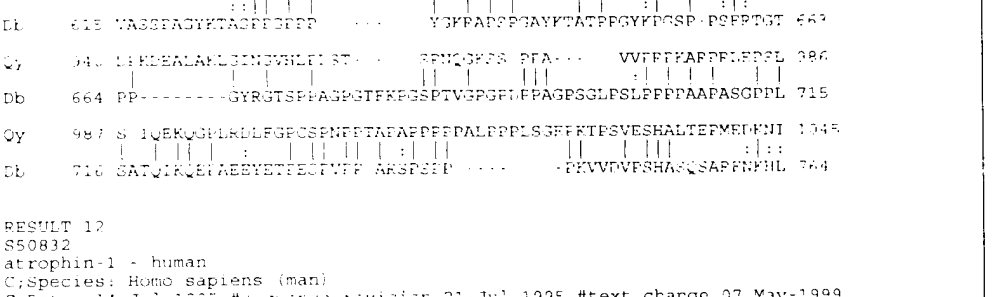
Qy 660 PAMGAVPTMPAMMVPPQPQHLVPGLPSPQPAIQVQNPTNLQAGILAGLMTTQAMSTGLPG 719
Db 360 PASSSAP- ---APPMRPPYSSDSSSSQAAASSSSSSSSSYA PPPIAPGAL 406

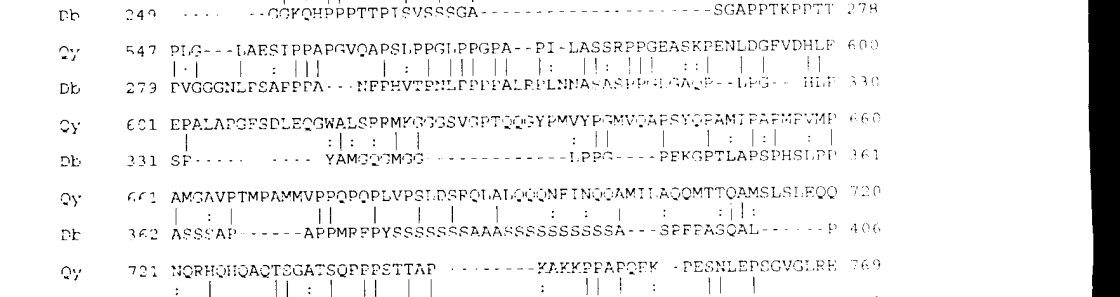
Qy 720 QNQRHQHQAQTSGATSQPPPSTTAP - WAPPPLAPLPP PPQNPPLPSNLP 769



RESULT 12
S50832
atrophin-1 - human
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C;Accession: S50832
R;Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.; Yam
Nature Genet. 8, 177-181, 1994
A;Title: Structure and expression of the gene responsible for the triplet repeat disorde
A;Reference number: S50832; MUID:95144175; PMID:7842016
A;Accession: S50832
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1184 <NAG>
A;Cross-references: EMBL:D31840
C;Genetics:
A;Gene: GDB:DRPLA; B37
A;Cross-references: GDB:270336; OMIM:125370
A;Map position: 12p13.31-12p13.3|12p-12p

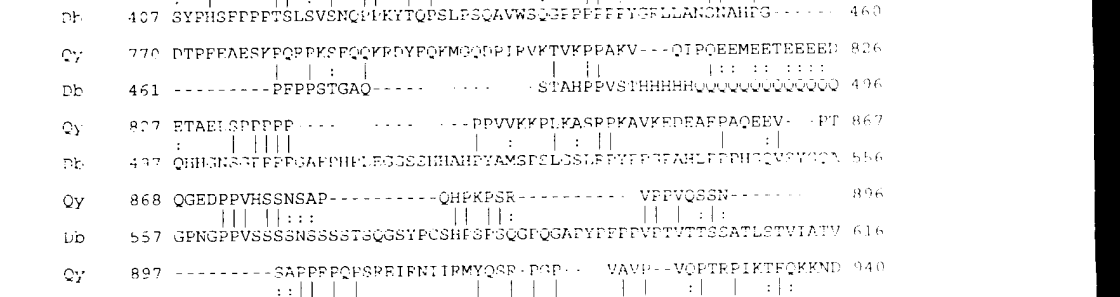
Query Match 4.1%; Score 286.5; DB 2; Length 1184;
Best Local Similarity 22.5%; Pred. No. 4.4e-06;
Matches 216; Conservative 91; Mismatches 313; Indels 340; Gaps 52;

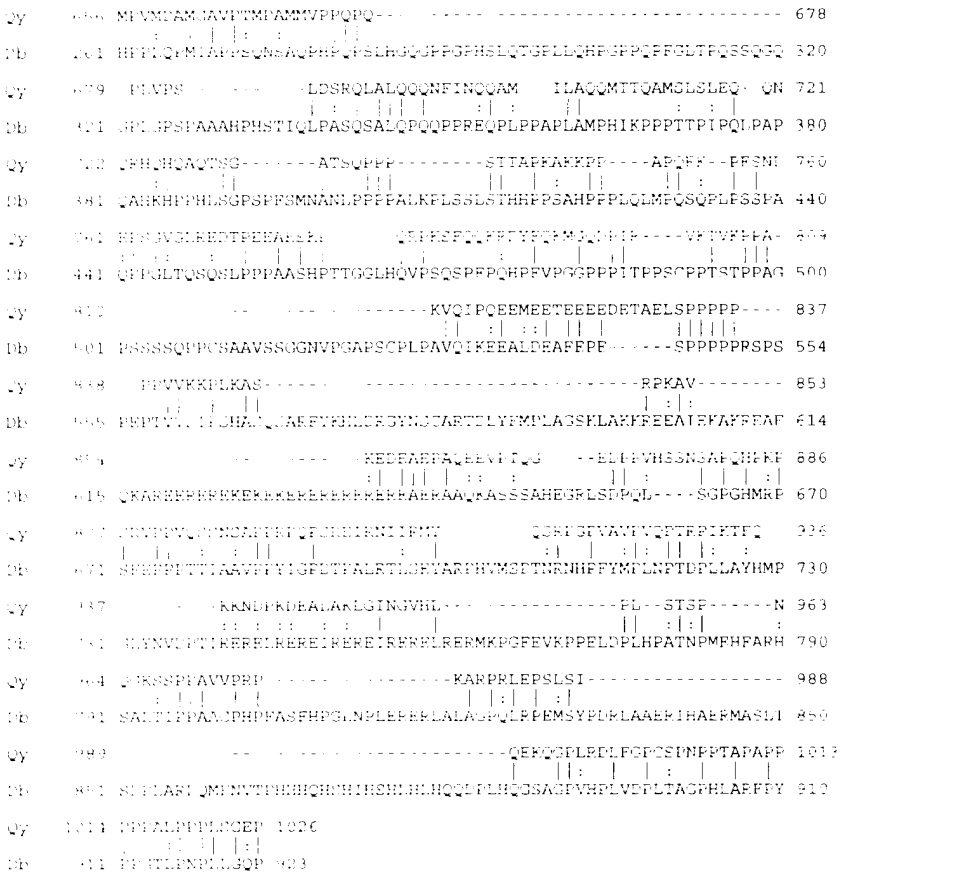




RESULT 13
T42731
atrophin-1 related protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42731
R;Khan, F.A.; Margolis, R.L.; Loev, S.L.; Sharp, A.H.; Li, S.H.; Ross, C.A.
submitted to the EMBL Data Library, December 1995
A;Description: cDNA sequence and expression of an atrophin-1 (DRPLA disease gene) relate
A;Reference number: Z22250
A;Accession: T42731
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1006 <KHA>
A;Cross-references: EMBL:U44091; NID:g1297319; PID:g1098103; PIDN:AAA88970.1
C;Genetics:
A;Gene: ARP

Query Match 4.1%; Score 286; DB 2; Length 1006;
Best Local Similarity 20.6%; Pred. No. 3.9e-06;
Matches 163; Conservative 61; Mismatches 205; Indels 364; Gaps 31;

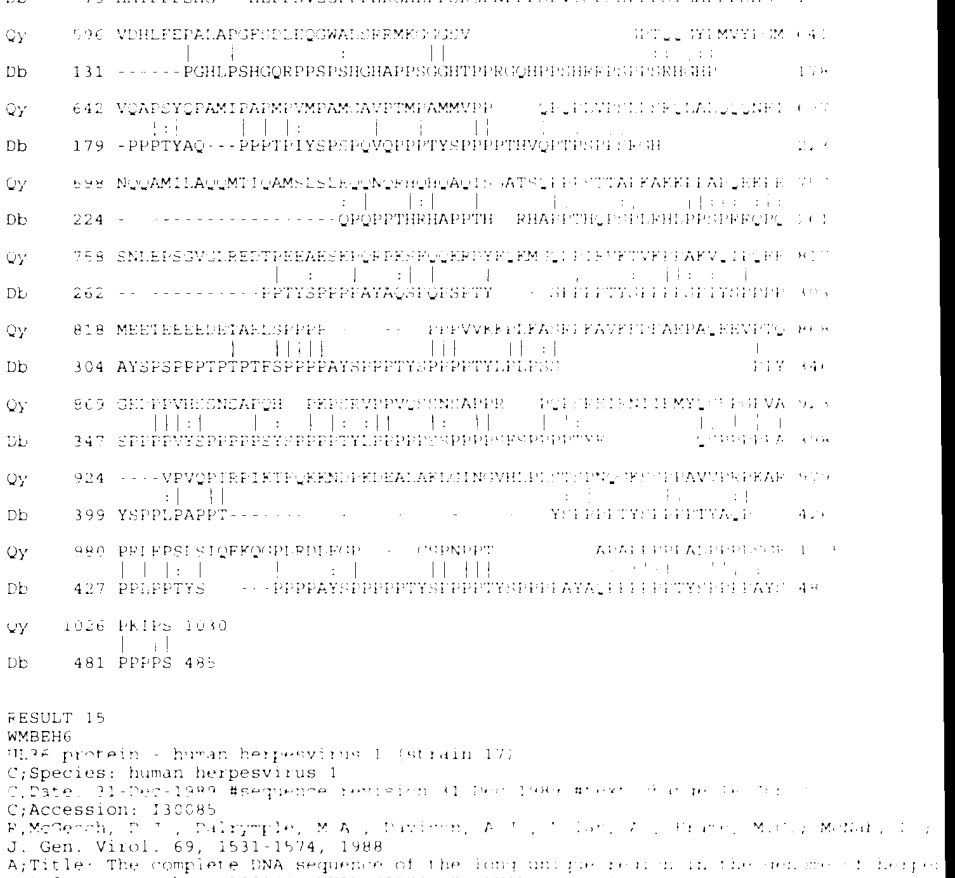




RESULT 14
S06733
hydroxyproline-rich glycoprotein precursor - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 16-Jul-1999
C;Accession: S06733
R;Keller, B.; Lamb, C.J.
Genes Dev. 3, 1639-1646, 1989
A;Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene
A;Reference number: S06733; MUID:90108263; PMID:2612999
A;Accession: S06733
A;Molecule type: DNA
A;Residues: 1-62 <KEL>
A;Cross-references: EMBL:X13664; NID:g13666; PIDN:CAA32030.1; PID:g17067
C;Superfamily: hydroxyproline-rich glycoprotein
C;Keywords: glycoprotein

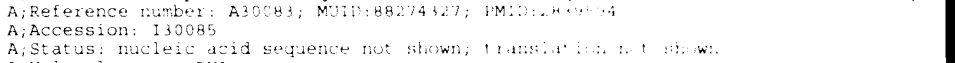
Query Match 4.0%; Score 281.5; DB 2; Length 620;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 144; Conservative 97; Mismatches 195; Indels 189; Gaps 23;

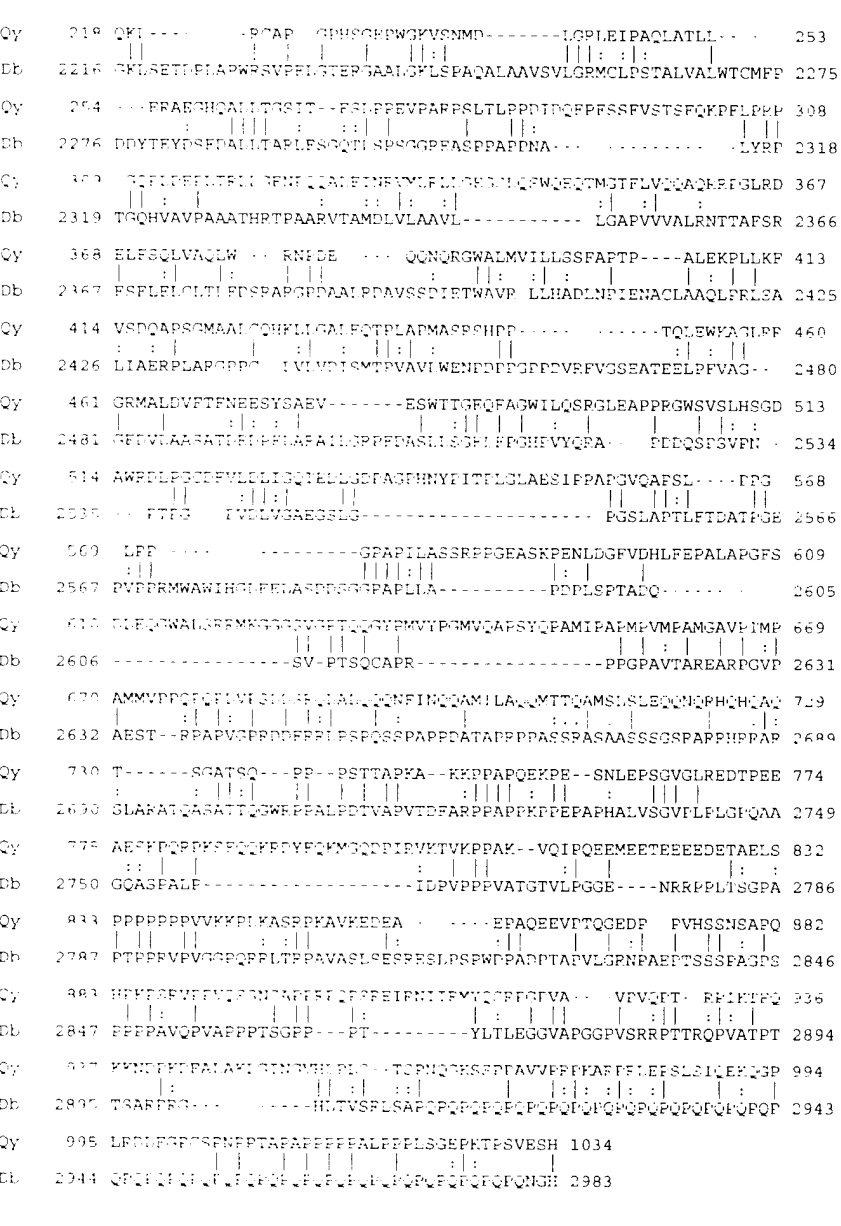




RESULT 15
WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)
C;Species: human herpesvirus 1
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 09-Jul-1999
C;Accession: I30085
R;McGeoch, D.J., Dalrymple, M.A., Davison, A.J., Dolan, A., Frame, M.C., McNab, D.
J. Gen. Virol. 69, 1531-1574, 1988
A;Title: The complete DNA sequence of the long unique region in the genome of herpes
A;Reference number: A30083; MUID:88274327; PMID:2839594
A;Accession: I30085
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3164 <MCG>
A;Cross-references: GB:X14112; NID:g1944586; PIDN:CAA32411.1; PID:g50546; SPTR:P0
C;Genetics:
A;Gene: UL36
C;Superfamily: varicella-zoster virus gene 22 protein

Query Match 4.0%; Score 280.5; DB 1; Length 3164;
Best Local Similarity 21.5%; Pred. No. 2.7e-05;
Matches 241; Conservative 104; Mismatches 466; Indels 107; Gaps 11;





Search completed: January 10, 2003, 02:24:16
Job time : 79.5262 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 22:12:30 ; Search time 38.6477 Seconds
(without alignments)
1434.857 Million cell updates/sec

Title: US-09-803-126-1
Perfect score: 6991
Sequence: 1 MGSIFQRAFPQAGTEQNFPT ... WQPPSKLGPAQEVPKTRRVK 1337

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

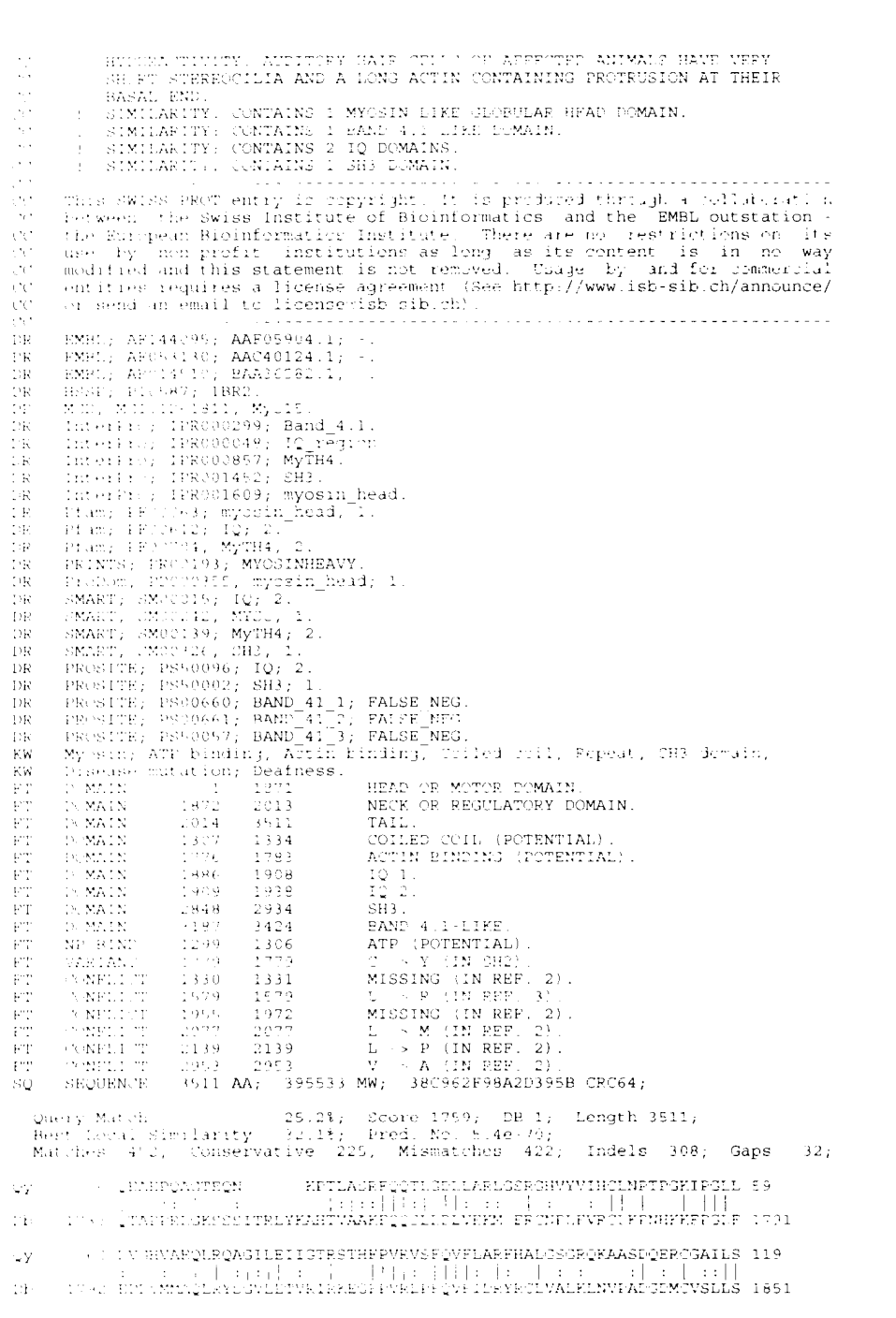
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1759	25.2	3511	1	MY15_MOUSE	Q9qzz4 mus musculu
2	1717	24.6	3530	1	MY15_HUMAN	Q9uku7 homo sapien
3	312.5	4.5	2215	1	MY7A_HUMAN	Q13402 homo sapien
4	303.5	4.3	2215	1	MY7A_MOUSE	P97479 mus musculu
5	298.5	4.3	1386	1	ZAP3_MOUSE	Q9r0i7 mus musculu
6	296.5	4.2	1193	1	DPPL_RAT	P54259 rattus norv
7	289.5	4.1	555	1	GP1_CHLRE	Q9fpg6 chlamydomon
8	287.5	4.1	1509	1	GRP1_HUMAN	Q9nzm4 homo sapien
9	286.5	4.1	1195	1	DPPL_HUMAN	P54259 homo sapien
10	281.5	4.0	620	1	EXIN_TOBAC	P13083 nicotiana t
11	280.5	4.0	3164	1	TEGU_HSV11	P10220 herpes simp
12	276.5	4.0	2142	1	BAT2_HUMAN	P48634 homo sapien
13	272.5	3.9	817	1	VRP1_YEAST	P37370 saccharomyc
14	268.5	3.8	746	1	DCAP_HUMAN	Q9frn5 homo sapien
15	261	3.7	1902	1	ZAP3_HUMAN	P49759 homo sapien
16	260.5	3.7	2161	1	SHK1_HUMAN	Q9y566 homo sapien
17	258.5	3.7	1815	1	SHK2_RAT	Q9jlh4 rattus norv
18	258.5	3.7	2162	1	SHK1_RAT	Q9wv49 rattus norv
19	257.5	3.7	1362	1	PRD4_HUMAN	O60885 homo sapien
20	253	3.6	2215	1	TPX2_HUMAN	Q9ulw0 homo sapien
21	252	3.6	1794	1	YAV1_SCHPO	Q10172 schizosacch
22	251.5	3.6	790	1	DCAP_MOUSE	Q924b0 mus musculu
23	248	3.5	2026	1	APP1_MOUSE	Q61329 mus musculu
24	245.5	3.5	865	1	CPN_DROME	Q02910 drosophila
25	241.5	3.5	3763	1	APP1_HUMAN	Q15911 homo sapien
26	241	3.4	1048	1	SPA4_RAT	Q63627 rattus norv
27	236.5	3.4	1210	1	AF4_HUMAN	P51825 homo sapien
28	233.5	3.3	267	1	EXTN_MAIZE	P14318 zea mays (m
29	233.5	3.3	938	1	EPN4_EBV	P03203 epstein-bar
30	233	3.3	503	1	WAIP_HUMAN	O43516 homo sapien
31	231	3.3	1087	1	NFH_MOUSE	P19246 mus musculu
32	231	3.3	2774	1	MAPA_RAT	P34926 rattus norv
33	230.5	3.3	283	1	EXTN_SORPI	P24152 sorghum bic
34	230	3.3	1567	1	FMN2_MOUSE	Q9jl04 mus musculu
35	229	3.3	1157	1	SPA4_HUMAN	Q96104 homo sapien
36	227	3.2	1018	1	RMW1_MYCTU	P63365 mycoplasma
37	225	3.2	623	1	LA17_YEAST	Q12446 saccharomyc
38	224	3.2	792	1	S3A1_HUMAN	P15498 homo sapien
39	223.5	3.2	1468	1	FMN1_MOUSE	Q61965 mus musculu
40	222.5	3.2	1980	1	MYOP_RAT	Q63259 rattus norv
41	221.5	3.2	879	1	Y563_HUMAN	P56563 homo sapien
42	221.5	3.2	1745	1	ZO1_MOUSE	P39447 mus musculu
43	220	3.1	426	1	EXLP_TOBAC	Q03211 nicotiana t
44	218	3.1	992	1	EBN6_EBV	P03204 epstein-bar
45	216	3.1	2442	1	CEP_HUMAN	Q92793 homo sapien

ALIGNMENTS

RESULT 1
MY15_MOUSE
ID MY15_MOUSE STANDARD; PRT; 3511 AA.
AC Q9QZZ4; O70395; Q9QWL6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin XV (Unconventional myosin-15).
GN MYO15A OR MYO15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20001762; PubMed=10553916;
RA Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
RA Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
RA Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,
RA Fridell R.A.;
RT "Characterization of the human and mouse unconventional myosin XV
RT genes responsible for hereditary deafness DFNB3 and shaker 2.";
RL Genomics 61:243-258(1999).
RN [2]
RP SEQUENCE OF 1169-2023 FROM N.A., AND VARIANT SER ILE 1774.
RC TISSUE=Embryo;
RX MEDLINE=98267310; PubMed=9603735;
RA Probst F.J., Fridell R.A., Raphael Y., Saunders T.L., Wang A.,
RA Liang Y., Morell R.J., Touchman J.W., Lyons R.H., Noben-Trauth K.,
RA Friedman T.B., Camper S.A.;
RT "Correction of deafness in shaker-2 mice by an unconventional myosin
RT in a BAC transgene.";
RL Science 280:1444-1447(1998).
RN [3]
RP SEQUENCE OF 1237-1823 FROM N.A., AND VARIANT SER ILE 1774.
RC STRAIN=C57BL/6;
RX MEDLINE=98369604; PubMed=9703981;
RA Wakabayashi Y., Takahashi Y., Kikkawa Y., Ikeda H., Mishima Y.,
RA Ushiki T., Yonekawa H., Hominami R.;
RT "A novel type of myosin encoded by the mouse deafness gene shaker-2.";
RL Biochem. Biophys. Res. Commun. 249:655-659(1998).
CC -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. MY15
CC SIMILARITY) MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: IN THE DEVELOPING INNER EAR, EXPRESSED IN
CC COCHLEA AND VESTIBULAR APPARATUS. EXPRESSION APPEARS TO BE
CC RESTRICTED TO COCHLEAR NEUROSENSORY CELLS AND INNER EPITHELIAL
CC LAYER OF MACULA SACCULA. ALSO EXPRESSED IN MACULA UTRICULI AND
CC CRISTAE AMPULLARIS OF THE SEMICIRCULAR CANALS. IN ADULT COCHLEAR
CC HAIR CELLS, HIGHEST EXPRESSION IN STEREOCILIA AND APICAL BODY.
CC -!- DISEASE: DEFECTS IN MYO15A ARE THE CAUSE OF SHAKER 2 (SH2), A
CC CONDITION CAUSING DEAFNESS, CIRCLING BEHAVIOR, HEAD-TOSSING AND


HYPERACTIVITY. AUDITORY HAIR CELLS OF AFFECTED ANIMALS HAVE VERY
SHORT STEREOCILIA AND A LONG ACTIN CONTAINING PROTRUSION AT THEIR
BASAL END.
SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
SIMILARITY: CONTAINS 1 BAND 4.1 LIKE DOMAIN.
SIMILARITY: CONTAINS 2 IQ DOMAINS.
SIMILARITY: CONTAINS 1 SH3 DOMAIN.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; AF144095; AAF05904.1; -.
DR EMBL; AF053130; AAC40124.1; -.
DR EMBL; AF161610; BAA36562.1; -.
DR HSSP; P15587; 1BR2.
DR MIM; 601317; MYO15.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00373; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF03176; MyTH4; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000315; myosin_head; 1.
DR SMART; SM00015; IQ; 2.
DR SMART; SM00242; MYSc; 1.
DR SMART; SM00139; MyTH4; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50096; IQ; 2.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS00660; BAND_4_1_1; FALSE_NEG.
DR PROSITE; PS00661; BAND_4_1_2; FALSE_NEG.
DR PROSITE; PS00057; BAND_4_1_3; FALSE_NEG.
KW Myosin; ATP binding; Actin binding; Coiled coil; Repeat, SH3 domain,
KW Disease mutation; Deafness.
FT DOMAIN 1 1871 HEAD OR MOTOR DOMAIN.
FT DOMAIN 1872 2013 NECK OR REGULATORY DOMAIN.
FT DOMAIN 2014 3511 TAIL.
FT DOMAIN 1307 1334 COILED COIL (POTENTIAL).
FT DOMAIN 1776 1783 ACTIN BINDING (POTENTIAL).
FT DOMAIN 1886 1908 IQ 1.
FT DOMAIN 1909 1938 IQ 2.
FT DOMAIN 2848 2934 SH3.
FT DOMAIN 2197 3424 BAND 4.1-LIKE.
FT NP BIND 1209 1306 ATP (POTENTIAL).
FT VARIANT 1778 1778 C -> Y (IN DH2).
FT CONFLICT 1330 1331 MISSING (IN REF. 2).
FT CONFLICT 1579 1579 L -> P (IN REF. 3).
FT CONFLICT 1969 1972 MISSING (IN REF. 2).
FT CONFLICT 2077 2077 L -> M (IN REF. 2).
FT CONFLICT 2139 2139 L -> P (IN REF. 2).
FT CONFLICT 2953 2953 V -> A (IN REF. 2).
SQ SEQUENCE 3511 AA; 395533 MW; 38C962F98A2D395B CRC64;
Query Match 25.2%; Score 1799; DB 1; Length 3511;
Best Local Similarity 72.1%; Pred. No. 5.4e-79;
Matches 413; Conservative 225; Mismatches 422; Indels 308; Gaps 32;

```
                :|  :|:|:   :  |||:|  ||||:||  |  |: ||:| ||| |::||:: :

Db   2733 LVTEPVPPAAISMAPESWPIYPSPLPPAMGSVGTGVQILAVSHTGIKLLQMVKGSKEASR 2792

Qy   1189 QLKTLQSVGVAFVLTVQCPRPSTLELSLKNFQLLILHTAWARAIKAMVDLPLSELRKDSDY 1248
          :|: ||:|:|:::|  |    .. || .| ||:||| :| |: :|:|| | || |||| |

Db   2793 PIPVLPAYEPACIIPVTMFSQNMLEPNLCNEPLILFSAPAQQVKTLVDTPILELKKDSDY 2852

Qy   1249 VIALRSYITDDNSLLSFHRGDLIRLLPVTALEP---------------------  1281
          |:|:|::::|  ||||:|:|:| | |    :|||

Db   2853 VVAVRNFLSEDPELLSFHVSDIIHI---QSLEPTPVGYSAGCVVPPKLVYLEPLPPGPD 2909

Qy   1282  GWQPGSASGRSGLEPDDVVQPAAAPD 1307
           |:|:|:  || | || ::|||||||
Db   2910 PGWRPGAVHGRVGRPPSELVQPAAAPD 2936


RESULT 2
MY15_HUMAN
ID   MY15_HUMAN      STANDARD;      PRT;  3530 AA.
AC   Q9UKN7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin XV (Unconventional myosin-15).
GN   MYO15A OR MYO15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX   MEDLINE=20021762; PubMed=10559926;
RA   Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
RA   Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
RA   Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,
RA   Fridell R.A.;
RT   "Characterization of the human and mouse unconventional myosin XV
RT   genes responsible for hereditary deafness DFNB3 and shaker 2.";
RL   Genomics 61:243-258(1999).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A., AND VARIANTS DFNB3 TYR-2111 AND PHE-2113.
RX   MEDLINE=98267311; PubMed=9603736;
RA   Wang A., Liang Y., Fridell R.A., Probst F.J., Wilcox E.R.,
RA   Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K.,
RA   Friedman T.B.;
RT   "Association of unconventional myosin MYO15 mutations with human
RT   nonsyndromic deafness DFNB3.";
RL   Science 280:1447-1451(1998).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. MY15
CC       SIMILARITY: MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
CC       THE ACTIN-RICH STEREOCILIA OF THE INNER EAR SENSORY HAIR CELLS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY. ALSO EXPRESSED
CC       AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,
CC       PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN, IN THE
CC       PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.
CC   -!- DISEASE: Defects in MYO15A are the cause of an autosomal recessive
CC       form of nonsyndromic deafness (DFNB3).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF144094; AAF05903.1; -.
DR   EMBL; AF051976; -; NOT_ANNOTATED_CDS.
DR   HSSP; P10587; 1BR2.
DR   Genew; HGNC:7594; MYO15A.
DR   MIM; 602666; -.
DR   MIM; 600316; -.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00660; BAND_4_1_1, FALSE_NEG.
DR   PROSITE; PS00661; BAND_4_1_2, FALSE_NEG.
DR   PROSITE; PS50057; BAND_4_1_3, 1.
DR   PROSITE; PS50096; IQ; 3.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Calmodulin-binding; Disease mutation; Deafness.
FT   DOMAIN        1   1887       HEAD OR MOTOR DOMAIN.
FT   DOMAIN     1888   2029       NECK OR REGULATORY DOMAIN.
FT   DOMAIN     2030   3530       TAIL.
FT   DOMAIN     1323   1350       COILED COIL (POTENTIAL).
FT   DOMAIN     1792   1799       ACTIN BINDING (POTENTIAL).
FT   DOMAIN     1902   1924       IQ 1.
FT   DOMAIN     1925   1954       IQ 2.
FT   DOMAIN     1955   1976       IQ 3.
FT   DOMAIN     2867   2953       SH3.
FT   DOMAIN     3206   3442       BAND 4.1-LIKE.
FT   NP_BIND    1315   1322       ATP (POTENTIAL).
FT   VARIANT    2111   2111       N   -> Y (IN DFNB3, FAMILY FROM BENGHALA).
FT                                /FTId=VAR_010303.
FT   VARIANT    2113   2113       I   -> F (IN DFNB3, INDIAN FAMILY).
FT                                /FTId=VAR_010304.
SQ   SEQUENCE   3530 AA,  405171 MW;  [illegible] CRC64;

  Query Match             24.6%;  Score 1717,  DB 1,  Length 3530;
  Best Local Similarity   31.6%;  Pred. No. 3.8e-68;
  Matches  447,  Conservative  213,  Mismatches  444,  Indels  310,  Gaps   32;

Qy       4 LPQEAEPQAGTEQ-------     NKITLASKFQQTLSELLARLGSSHVYVIHCLNPTP 52
           ||    ||| .:               |.|.:|||:| ||:.::  |:   : || |
Db    1742 IFSSHAPQAAPQPLGKSSSVTFIYFAHTVAAPFQQSLLPLVEPM-ERCNILEMKCLRIMH 1800

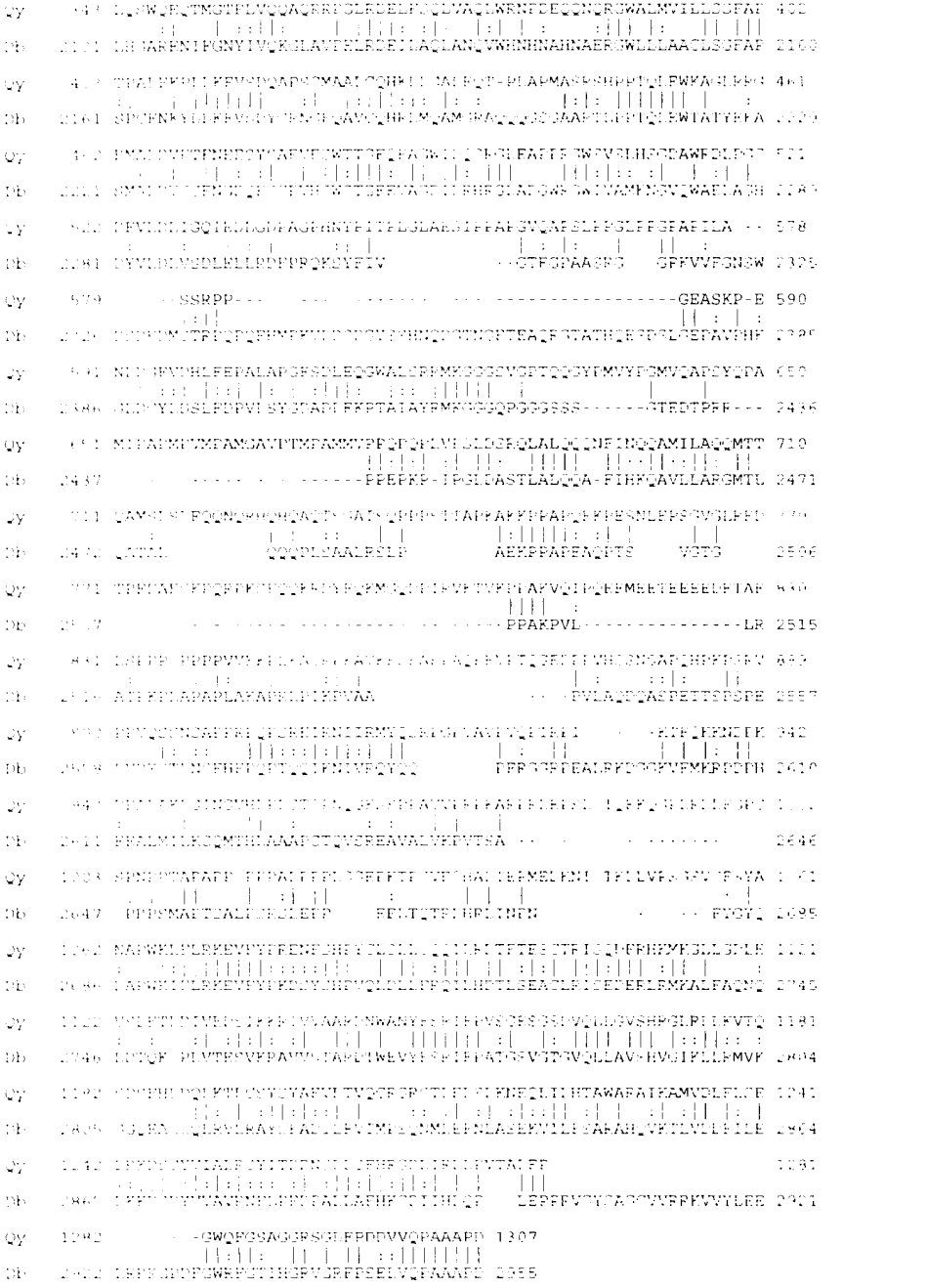
Qy      53 GKIPGLIDVGHVAEQLPGAIIEIISTFSTHFPVPVEFQVSLAEFHALESSPQPAACFQE 112
           | ||| :   |  ||| :|.|| :  |   |||| || |: |:  |  : :    :: :
Db    1801 FKEPGLFPDVVMAADLFYSGVIETVPIFFESFPVELPFQGPICPYCCLVALPHDLPANCP 1860

Qy     113 PQSATLSEVLQAESPIYHLGVTQVLLQEQSWPQLEQLWAIPRSQALLILHRGLRACITRQ 172
           |   ||      :  ||: :|| .  : || .    : |||| | ||   ::
Db    1861 MPVSVLSFLCKVMINMYEVGVSELPLKEHLTYQILESMPEHVLNLAADTLPFQLPSFIRF 1920

Qy     173 PLSLLPP----MQAPVPGLIAFRPYLQPSALGQLNTIL   LVAFILSSRPKLKI AP 224
           | | |      :|:| ||  ||:|| | | .|.  ...   :  |  |: | : ||
Db    1921 RPRSIPHPTILLQSPAPGYLAPQPYQQMPFSLVFPFCLVHAYVSPRFTLPIGAPWPQLVL 1980

Qy     225 GP--HGGEPWGKVSNMDLGPLETPAQLATLLEPAFSHQALQTGSITELPIEVPARLLL 282
           |    |  . .| ||:|||| ||:     |          :    | : | | :|
Db    1981 GALLWRQEELSKPEVVAVGHLFVPAELAGLQAVAGLGLAQVPQVAPVPTPFLQAPPPVL 2040

Qy     283 IPPDIDQPFPSSPVSTSPQKPPLPSPDQPLGPPLTPLGSPNPLALPINPVMLPLLGPGL 342
           || ||: :| : ||   |:.|       ||  |||:|  |. .:|. | ...|| :|:
Db    2041 LPLGINNYTPMAKFVQCHPKEPAPGMLTVPLRTPLTQLFAEHHAEAVSIPKLILAPMGDPH 2100
```



```
RESULT 3
MY7A_HUMAN
ID   MY7A_HUMAN     STANDARD;      PRT;  2215 AA.
AC   Q13402; Q92821; Q92822; Q13401; P78427; Q14784;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin VIIa.
GN   MY7A OR USH1B.
OS   Homo sapiens (Human).
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1, 5, 6 AND 7), AND DEVELOPMENTAL
RP   EXPRESSION.
RC   TISSUE=Retina;
RX   MEDLINE=96194949; PubMed=8622919;
RA   Weil D., Levy G., Sahly I., Levi-Acobas F., Blanchard S.,
RA   El-Amraoui A., Crozet F., Philippe H., Abitbol M., Petit C.;
RT   "Human myosin VIIA responsible for the Usher 1B syndrome: a predicted
RT   membrane-associated motor protein expressed in developing sensory
RT   epithelia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3232-3237(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 4).
RC   TISSUE=Testis;
RX   MEDLINE=97038686; PubMed=8884267;
RA   Chen Z.-Y., Hasson T., Kelley P.M., Schwender B.J., Schwartz M.F.,
RA   Ramakrishnan M., Kimberling W.J., Mooseker M.S., Corey D.P.;
RT   "Molecular cloning and domain structure of human myosin-VIIa, the gene
RT   product defective in Usher syndrome 1B.";
RL   Genomics 36:440-448(1996).
RN   [3]
RP   SEQUENCE OF 166-196 FROM N.A.
RC   TISSUE=Epithelium, Leukocyte, and Liver;
RX   MEDLINE=94344815; PubMed=8022818;
RA   Bement W.M., Hasson T., Wirth J.A., Cheney R.E., Mooseker M.S.;
RT   "Identification and overlapping expression of multiple unconventional
RT   myosin genes in vertebrate cell types.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6549-6553(1994).
RN   [4]
RP   SEQUENCE OF 1-1078 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=96-----; PubMed=7568024;
RA   Hasson T., Heintzelman M.B., Santos-Sacchi J., Corey D.P.,
RA   Mooseker M.S.;
RT   "Expression in cochlea and retina of myosin VIIa, the gene product
RT   defective in Usher syndrome type 1B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:9815-9819(1995).
RN   [5]
RP   SEQUENCE OF 96-564 FROM N.A. (ISOFORM 1), AND VARIANTS USH1B.
RC   TISSUE=Retina;
RX   MEDLINE=95174880; PubMed=7870171;
RA   Weil D., Blanchard S., Kaplan J., Guilford P., Gibson F.,
RA   Walsh J., Mburu P., Varela A., Levilliers J., Weston M.D.,
RA   Kelley P.M., Kimberling W.J., Wagenaar M., Levi-Acobas F.,
RA   Larget-Piet D., Munnich A., Steel K.P., Brown S.D.M., Petit C.;
RT   "Defective myosin VIIA gene responsible for Usher syndrome type 1B.";
RL   Nature 374:60-61(1995).
RN   [6]
RP   SEQUENCE OF 79-578 FROM N.A.
RX   MEDLINE=97224487; PubMed=9070921;
RA   Kelley P.M., Weston M.D., Chen Z.-Y., Orten D.J., Hasson T.,
RA   Overbeck L.D., Pieke J., Talmadge C.B., Isaacson P.,
RA   Corey D.P., Sumegi J., Kimberling W.J.;
RT   "The genomic structure of the gene [illegible] Usher syndrome 1B
RT   (MYO7A).";
RL   Genomics 40:73-79(1997).
RN   [7]
RP   DEVELOPMENTAL EXPRESSION.
RX   MEDLINE=96440426; PubMed=8842737;
RA   El-Amraoui A., Sahly I., Picaud S., Sahel J., Abitbol M., Petit C.;
```

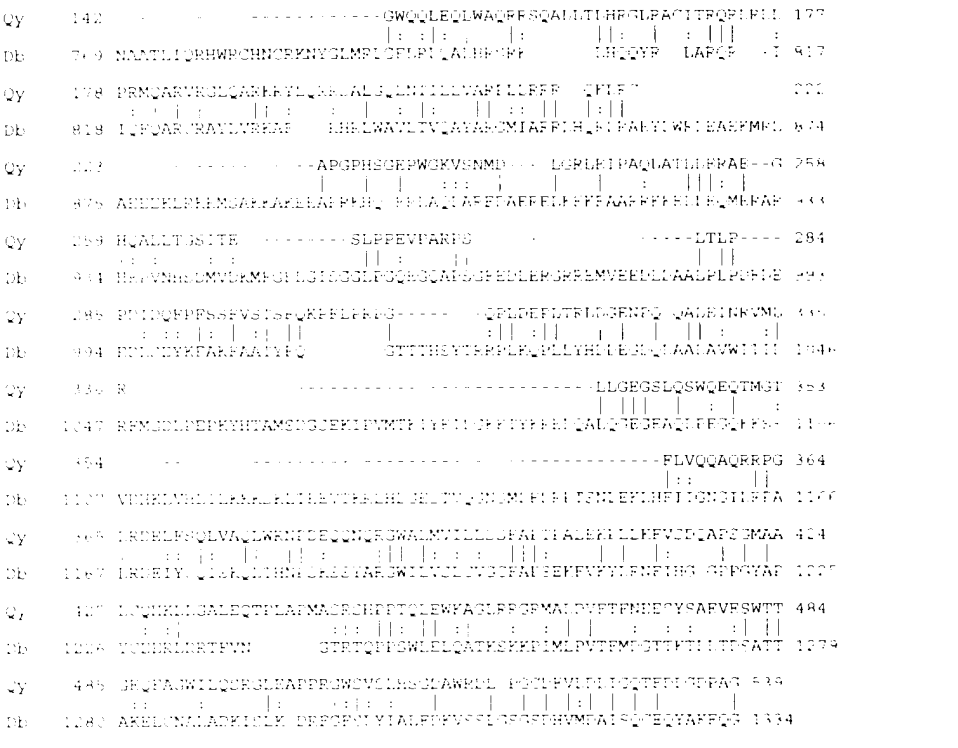
```
PT   "Human Usher 1B/mouse shaker-1: the retinal phenotype discrepancy
PT   explained by the presence/absence of myosin VIIA in the photoreceptor
RT   cells.";
RL   Hum. Mol. Genet. 5:1171-1178(1996).
RN   [8]
RP   VARIANTS USH1B H-212, C-212, H-332, Q-468 INS; L-503 AND Q-450.
RX   MEDLINE=97055890; PubMed=9000236;
RA   Weston M.D., Kelley P.M., Overbeck L.D., Wagenaar M., Orten D.J.,
RA   Hasson T., Chen Z.-Y., Corey D.P., Mooseker M.S., Sumegi J.,
RA   Cremers C., Moeller C., Jacobson S.G., Gorin M.B., Kimberling W.J.;
PT   "Myosin VIIA mutation screening in 189 Usher syndrome type 1
RT   patients.";
RL   Am. J. Hum. Genet. 59:1074-1083(1996).
RN   [9]
RP   VARIANTS USH1B ARG-214, ASP-237 AND THR-806, AND POLYMORPHISMS.
RX   MEDLINE=98019437; PubMed=9392681;
RA   Adato A., Weil D., Kalinski H., Pel-Or Y., Ayadi H., Petit C.,
RA   Korostishevsky M., Bonne-Tamir B.;
RT   "Mutation-profile of all 49 exons of the human myosin VIIA gene, and
RT   haplotype analysis, in Usher 1B families from diverse origins.";
RL   Am. J. Hum. Genet. 61:813-821(1997).
RN   [10]
RP   VARIANTS USH1B ARG-05, PRO-906 AND GLN-2137, AND POLYMORPHISMS.
RX   MEDLINE=97156018; PubMed=9009278;
RA   Levy G., Levi-Acobas F., Blanchard S., Gerber S., Larget-Piet D.,
RA   Chenal V., Liu X.-Z., Newton V., Steel K.P., Brown S.D.M., Munnich A.,
RA   Kaplan J., Petit C., Weil D.;
RT   "Myosin VIIA gene heterogeneity of the mutations responsible for
RT   Usher syndrome type IB.";
RL   Hum. Mol. Genet. 6:111-116(1997).
RN   [11]
RP   VARIANT DFNB2 PRO-244.
RX   MEDLINE=97319827; PubMed=9171833;
RA   Liu X.-Z., Walsh J., Mburu P., Kendrick-Jones J., Cope M.J.,
RA   Steel K.P., Brown S.D.M.;
RT   "Mutations in the myosin VIIA gene cause non-syndromic recessive
RT   deafness.";
RL   Nat. Genet. 16:188-190(1997).
RN   [12]
RP   VARIANT DFNB2 ILE-503.
RX   MEDLINE=97315844; PubMed=9171833;
RA   Weil D., Kussel P., Blanchard S., Levy G., Levi-Acobas F., Drira M.,
RA   Ayadi H., Petit C.;
RT   "The autosomal recessive isolated deafness, DFNB2, and the Usher 1B
RT   syndrome are allelic defects of the myosin-VIIA gene.";
RL   Nat. Genet. 16:191-193(1997).
RN   [13]
RP   VARIANT DFNA11 ARG-ALA--LYS 888 DEL.
RX   MEDLINE=98216494; PubMed=9554094;
RA   Liu X.-Z., Walsh J., Tamagawa Y., Kitamura K., Nishizawa M.,
RA   Steel K.P., Brown S.D.M.;
RT   "Autosomal dominant non-syndromic deafness caused by a mutation in the
RT   myosin VIIA gene.";
RL   Nat. Genet. 17:268-269(1997).
RN   [14]
RP   VARIANTS USH1B PRO-651 AND GLN-1602.
RX   MEDLINE=98386090; PubMed=9718356;
RA   Liu X.-Z., Hope C., Walsh J., Newton V., Ke X.-M., Liang C.-Y., Xu L.-P.,
RA   Zhou J.-M., Tamagawa Y., Steel K.P., Bundey S., Brown S.D.M.;
RT   "Mutations in the myosin VIIA gene cause a wide phenotypic spectrum,
RT   including atypical Usher syndrome.";
RL   Am. J. Hum. Genet. 63:909-917(1998).
RN   [15]
RP   VARIANT USH1B PRO-1787.
RX   MEDLINE=99294608; PubMed=10364543;
RA   Adato A., Kalinski H., Weil D., Chaib H., Korostishevsky M.,
RA   Bonne-Tamir B.;
RT   "Possible interaction between USH1B and USH3 gene products as implied
RT   by apparent digenic deafness inheritance.";
RL   Am. J. Hum. Genet. 65:261-265(1999).
RN   [16]
RP   VARIANTS USH1B, AND POLYMORPHISMS.
RX   MEDLINE=99140257; PubMed=10094549;
RA   Janecke A.R., Weins M., Sadeghi M., Grundmann K., Apfelstedt-Sylla E.,
RA   Zrenner E., Rosenberg T., Gal A.;
PT   "Twelve novel myosin VIIA mutations in 34 patients with Usher syndrome
PT   type I: confirmation of genetic heterogeneity.";
RL   Hum. Mutat. 13:133-140(1999).
RN   [17]
RP   VARIANTS USH1B LYS-1170 AND CYS-1719.
RX   MEDLINE=99374392; PubMed=10447383;
RA   Cuevas J.M., Espinos C., Millan J.M., Sanchez F., Trujillo M.J.,
RA   Ayuso C., Beneyto M., Najera C.;
PT   "Identification of three novel mutations in the MYO7A gene.";
RL   Hum. Mutat. 14:181-181(1999).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN
CC       RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RIBBON
CC       SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTORS
CC       DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE POLARITY OF STEREOCILIA
CC       DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE. INVOLVED IN HAIR-CELL
CC       VESICLE TRAFFICKING OF AMINOGLYCOSIDES, WHICH ARE KNOWN TO INDUCE
CC       OTOTOXICITY (BY SIMILARITY).
CC   -!- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO-HEADED MOLECULE THROUGH THE
CC       FORMATION OF A COILED-COIL ROD.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PERIPHERAL). IN THE PHOTORECEPTOR
CC       CELLS, MAINLY LOCALIZED IN THE INNER AND BASE OF OUTER SEGMENTS AS
CC       WELL AS IN THE SYNAPTIC ENDING REGION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS, 1 (SHOWN HERE), 2, 3,
CC       4, 5, 6 AND 7, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    !  TISSUE SPECIFICITY: EXPRESSED IN THE PIGMENT EPITHELIUM AND THE
CC       PHOTORECEPTOR CELLS OF THE RETINA. ALSO FOUND IN KIDNEY, LIVER,
CC       TESTIS, COCHLEA, LYMPHOCYTES. NOT EXPRESSED IN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: DETECTED IN OPTIC CUP IN 5.5 WEEKS OLD
CC       EMBRYOS. EXPRESSED IN RETINAL PIGMENT EPITHELIUM, COCHLEAR AND
CC       VESTIBULAR NEUROEPITHELIA, AND OLFACTORY EPITHELIUM AT 9 WEEKS. AT
CC       10 WEEKS, PRESENT IN BOTH PIGMENT EPITHELIUM AND PHOTORECEPTOR
CC       CELLS. AT 24-28 WEEKS, EXPRESSION IN PIGMENT EPITHELIUM AND
CC       PHOTORECEPTOR CELLS INCREASES. PRESENT IN PIGMENT EPITHELIUM AND
CC       PHOTORECEPTOR CELLS IN ADULT.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF USHER SYNDROME TYPE 1B
CC       (USH1B), ALSO KNOWN AS USHER SYNDROME TYPE 1 ARABIAN VARIETY. USHER
CC       SYNDROME TYPE 1 IS AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY
CC       PROFOUND CONGENITAL SENSORINEURAL DEAFNESS, CONSTANT VESTIBULAR
CC       DYSFUNCTION AND PREPUBERTAL ONSET OF PROGRESSIVE RETINITIS
CC       PIGMENTOSA LEADING TO BLINDNESS. USHER SYNDROME IS THE MOST COMMON
CC       CAUSE OF COMBINED DEAFNESS AND BLINDNESS IN DEVELOPED COUNTRIES.
CC       USHB1 APPEARS AS A PRIMARY CYTOSKELETAL PROTEIN DEFECT.
CC    !  DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB2) MAPPED WITH
CC       VARIABLE VESTIBULAR DYSFUNCTION AND VARIABLE AGE OF ONSET.
CC   -!- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA11) WITH ONSET
CC       AFTER COMPLETE SPEECH ACQUISITION AND SUBSEQUENT GRADUAL
CC       PROGRESSION.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 IQ DOMAINS.

  Query Match             4.5%;  Score 212.5;  DB 1;  Length 2215;
  Best Local Similarity   21.0%;  Prod. No. 1.2e-06;
  Matches  163;  Conservative   94;  Mismatches  246;  Indels  273;  Gaps   27;

Qy       4 LPQEAADPQAGTEQQR--PTLASPPVQTLDLLAALGSSPGHVYYIHCLNPTPGKIPGLLDV 61
           .|| |:   | |  |  |||.|.|...|  |:  ||:    : : |: |   | | | |
Db     534 IPQ ADVAMGARTPPPSPTLSSQPFPSIKLLMPEIGSA-C-LPPVPTPPNEPPPTMLPPR 649

Qy      62 GHVAPQLPQAGILFIGTPSTHEPVPVSEQVPLAPPHALGSGPQPAASPQP---PCGAIL 118
           ||| ·|··| |  |   :|·| || |· |·  |  | : |    :    | :
Db     650 HLCVPQLPYSGMMETIPPPAGVEPPYSPVEPVEPYPVLLPGVKPAYPQGDLPGTPQPMA 709

Qy     119 SEVLGAESPLYHLGVTQVLLQEQ---------------------------------- 141
           |||     : :| |:: |::
Db     710 EAVLGTHED-WQIGKTPIFLPGHHLMLLEVPPLPAITEPVILPVIS-PVIS-PPIP ND:PIP 768
```



```
RESULT 4
MYO7A_MOUSE
ID   MYO7A_MOUSE     STANDARD;      PRT;  2215 AA.
AC   P97479;
DT   16-OCT-1997 (Rel. 35, Created)
DT   16-OCT-1997 (Rel. 35, Last sequence update)
DT   15-JUL-2001 (Rel. 39, Last annotation update)
DE   Myosin VIIA.
GN   MYO7A OR MY07.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX   NCBI TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97174881; PubMed=9070172;
RA   Gibson F., Walsh J., Mburu P., Varela A., Brown K.A., Antonio M.,
RA   Beisel K.W., Steel K.P., Brown S.D.M.;
RT   "A type VII myosin encoded by the mouse deafness gene shaker 1.";
RL   Nature 374:62-64(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98343888; PubMed=9680294;
RA   Mburu P., Liu X.-Z., Walsh J., Saw D. Jr., Cope M.J., Gibson F.,
RA   Kendrick-Jones J., Steel K.P., Brown S.D.M.;
RT   "Mutation analysis of the mouse myosin VIIA deafness gene.",
RL   Genes Funct. 1:191-203(1997).
CC   -!- FUNCTION: MYOSINS ARE ACTIN BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS DRIVE INTRACELLULAR MOVEMENTS
CC       WITH HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN
CC       HUMAN, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RIBBON
CC       SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTOR
CC       DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE POLARITY OF STEREOCILIA
CC       DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE BY SIMILARITY.
CC       INVOLVED IN HAIR-CELL VESICLE TRAFFICKING OF AMINOGLYCOSIDES,
CC       WHICH ARE KNOWN TO INDUCE OTOTOXICITY.
CC   -!- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE THROUGH THE
CC       FORMATION OF A COILED COIL BY SIMILARITY.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE RETINA, COCHLEA, KIDNEY AND
CC       LIVER.
CC   -!- DEVELOPMENTAL STAGE: IN THE INNER EAR OF THE 16.5 DAY (D) EMBRYO,
CC       EXPRESSED ONLY IN THE COCHLEAR AND VESTIBULAR SENSORY HAIR CELLS.
CC       IN ADDITION, EXPRESSION ALSO OCCURS IN THE EPITHELIAL CELLS OF THE
CC       SMALL INTESTINE, HEPATOCYTES, AND CHOROIDAL PLEXUS.
CC   -!- DISEASE: EFFECTS IN MYO7A ARE THE CAUSE OF THE SHAKER 1 (SH1)
CC       PHENOTYPE WHICH AFFECTS ONLY THE INNER EAR. SH1 HOMOZYGOTE
CC       MUTANTS SHOW HYPERACTIVITY, HEAD TOSSING AND CIRCLING DUE TO
CC       VESTIBULAR DYSFUNCTION, TOGETHER WITH TYPICAL NECK EPITHELIAL TYPE
CC       COCHLEAR EFFECTS INVOLVING DYSFUNCTION AND PROGRESSIVE
CC       DEGENERATION OF THE ORGAN OF CORTI.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN HEAD-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FAN 4.1 LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U81453; AAB40708.1; -.
DR   HSSP; P08799; 1MND.
DR   MGD; MGI:104510; Myo7a.
DR   InterPro; IPR000299; Band 4.1.
DR   InterPro; IPR000048; IQ region.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin head.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00784; MyTH4; 2.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000066; SH3; 1.
DR   ProDom; PD000855; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00660; BAND_4_1_1; FALSE_NEG.
DR   PROSITE; PS00661; BAND_4_1_2; FALSE_NEG.
DR   PROSITE; PS00057; BAND_4_1_3; 2.
DR   PROSITE; PS50096; IQ; 5.
DR   PROSITE; PS50002; SH3; 1.
KW   Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Calmodulin-binding; Disease mutation; Deafness.
FT   DOMAIN          1    729       MYOSIN HEAD-LIKE.
FT   DOMAIN        745    765       IQ 1.
FT   DOMAIN        768    788       IQ 2.
FT   DOMAIN        791    811       IQ 3.
FT   DOMAIN        814    834       IQ 4.
FT   DOMAIN        837    857       IQ 5.
FT   DOMAIN        918    935       COILED COIL (BY SIMILARITY).
FT   DOMAIN       1255   1441       BAND 4.1-LIKE 1.
FT   DOMAIN       1603   1670       SH3.
FT   DOMAIN       1869   2139       BAND 4.1-LIKE 2.
FT   NP_BIND       159    166       ATP (BY SIMILARITY).
FT   DOMAIN        632    639       ACTIN-BINDING (BY SIMILARITY).
FT   VARIANT       241    241       R -> P (IN SH1).
```

```
FT   VARIANT      502    502       R  > P (IN SH-1).
SQ   SEQUENCE   2215 AA,  254847 MW,  34740DD1941AB2E0 CRC64;

  Query Match             4.3%;  Score 238.5;  DB 1,  Length 2215;
  Best Local Similarity   21.3%;  Pred. No. 3e-06;
  Matches  167;  Conservative   92;  Mismatches  236;  Indels  289;  Gaps   29;

Qy       4 LPQRAPPQAGTFQNP--PTLASPPQQTLGPLLAPLGSPGHVYVTHCLNPTPGKIPGLLDV 61
           :|| |:   | |  |  |||:|:|:::|  |:  ||:    . .|. |  | | | |
Db     570 TP  AVAMAARTPPFSPTLGSLPPPSGPLLGMPTLGA-GQPPPVPGIPPNSPKPPMLPGP 628

Qy      62 GHVASQLPQAGILPTIGTRSTHPPVPVSPQVPLAPPHALGSGPQFAASDQE---PGGAIL 118
           ||| :|:-| |  |   :|:| || |  |:  |  | :|    :   |  :
Db     629 HL VPGIPPV MMPGTIPIPHA PPIPVSPVPPGPPVPVLLPGVPPAYPGGLPGTPQPMA 688

Qy     119 SEVLGAESPLYHLGVTQVLLQEQ                                 --- 141
           |||      : :| |:: |::
Db     689 PAVITTHTT-WQIGPTPIPLPTHHPMLLEVPPIPAITDPVILLQPVIPGPKDPSNPLPLP 748

Qy     142 ------------ ----------GWQQLPQLWAQPPSQALLTLHPGLPAGITPQPLPLLL 177
                                  |:|: |   |:     ||:  |  : |||   :
Db     749 CAATLIGPHNPGHHIGPPNYELIPLDPLPLLALHPSPP   --- LHPGVP--IAPGP---I 817

Qy     178 PPMQAPVPPIQAPPPYLQPPSALGQLNTIILVAPPLLPPP-------------------- 217
           |||  |    || :   |  |  : |.   || .. ||
Db     818 IPPQARGPAYLVPPAP   PHPLWAVITVQAYAPGMIAPPLHPPLPVPYQPPLEAEPMPL 874

Qy     218     QKLRGAPGPHEGEPWGKVGNMDLGR- LEIPAQLA-----------------TILE 264
               ||||       |  :  |   |  ||||                      |||
Db     875 AEEEKIP       PEMGAPPAPEPAEPPHQEPIAQLAPPDAEPPLPPPEEAPPPKKELE 927

Qy     265 PAE--GHQALLIGSIIE----------SLPPEVPARPS-------------------LT 282
           : |   |: :   : :           |||  :   ||                    |
Db     928 QMEPAPHEIPNGLMVLPMPGPLGTGSGLPQEGGAPGPPPMVPEPQVPAALP 980

Qy     283 LP----PPIPQPPPSPVPTGPLPPPLPPGS----     QPLPPPLTPLPSENPQ-QALE 329
           ||     |: :: |: | :| ||        |          :|| :||   | |   | ||
Db     988 LPDPGLPPPLPPYPPAPPAAIYPP        GTTIPGYTPPPLPLQPLIYHPDPGPQIAALA 1040

Qy     330 INPVMLPLLS --        EGS ---          ----------LQSWQEQIMGTPLV 366
           :  :|| :|             :||                    ||: | :  |  .
Db    1041 YWITILPPMPPPLPPPYYHTAMPPPGPEKIPVMTPIYETLGPPTYPPPELQAALQQP---GEPQL 1097

Qy     367 QQAQR---------------------------------------------------- 371
           : |:
Db    1098 PPQQPPPSVPHPIVHPTIPPPGPITPPVTPPLNPHPPTVQSNPMLEPPPTSNLPPIHPII 1158

Qy     372 ----RPSLPLPLPQLVALLWRNPLPQQNQPGWALMVILLSSPAPTPALPPLLPFVSP 426
               || ||||:: |:  ||  || .  ||| |. . :  |||.      | |  |.
Db    1159 SNSILPPALPCPITQQIPHLTHNPSPSSYAPGWILVSLCVGCPAPGPPPVPVPLPNPIHG 1219

Qy     427 QAPSGMAAAIQHPLLGALEQIPLAPMASPSHPPTQLPWPAGLPPGPMALDVPTPNEESYS 486
           | | |  |. .|          .|. ||. || .|   :  :| |  .  : :
Db    1220  GPPPYAPY PPPLPPTPVN   --  PTPTQIPGWLPLQATPSPPPIMLIVIPMPGTTPT 1278

Qy     487 AEVPSWTTGPLPAGWILQGPGLPAPPPGWGVSLHGGPAWPPL-PGCGPVLPLIGLTPTGG 545
           :| ||  .    .     |.   |.|. .  |   |  | | |:| | | |
Db    1279 LLTPSATTAPPLGNALAPPISLPPPPPGPGLPLYIALPDPVSSLGSGSDHVMDAISPQPLYA 1338

Qy     546 DIAG 549
           |
Db    1339 KEQG 1342


RESULT 5
ZAP3_MOUSE
ID   ZAP3_MOUSE      STANDARD;      PRT;   1380 AA.
AC   Q9R0I7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear protein ZAP3.
GN   ZAP3 OR ZAP.

OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal liver;
RA   Misawa K., Nosaka T., Kitamura T.;
RT   "A huge nuclear protein rich in proline similar to human hypothetical
RT   protein zap3 and zap113.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collab
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB033168; BAA85180.1; -.
DR   MGD; MGI:1926195; Zap3.
KW   Nuclear protein.
FT   DOMAIN       15    204       PRO-RICH.
FT   DOMAIN      355    473       GLN-RICH.
FT   DOMAIN      925   1012       ARG-RICH.
SQ   SEQUENCE   1386 AA,  155130 MW;  D862B9818BD221DF CRC64;

  Query Match             4.3%;  Score 238.5;  DB 1,  Length 1386;
  Best Local Similarity   19.5%;  Pred. No. 3e-06;
  Matches  241;  Conservative  116;  Mismatches  368;  Indels  509;  Gaps   60;

Qy     269 EGLPPEVPAPPSLTLPPGIQP     PPPGPVGTSPQPPPLPPPPQPLPPPLTPLPGPNP 324
           || |   |  |   :||          |  |:   |  :|| | ||           |
Db     145 ESPPEPPPVPPGPYMPPSQPYMPPPQPPPPYYPPPSSAQPYLP-PAQP        PPPPP 196

Qy     325 QQALEINPVMLPLLGPGSLQSWQEQTMGTPLVQQAQPPPQLPPPLPLPGPLVAQIWPPP  382
           |          |     |: | :      :||     | :::   |  |    :|
Db     197 Q---------- LPPPPSIPSGNPTA-----PQQPPLPSGAPNPTAPQPQPAPPPDPGTM 239

Qy     383  EQQNLPGW---ALMVILLSSPAPTPALPPLLP      PPVGPQAPGPMAALQPHPLLAT 432
            |: | |     |  :  |  : | |:      : : ::||   |:       |
Db     240 TPQEQQQYWYPQHILPQQPTPVHLPGHPPGPLVTAPDVPPPIPPPAPGPAAQQVAPPI-  297

Qy     433 ALEQTPLAPMASPSHPPTQLE      ------------------ WPAGLPPGP--  462
           | |: || |     : ||  |                        |:|     :
Db     298 AAEPPPLPPPNPPAPPPLSPEPPQSEDSPDSPDSPPDAPPPKQIPAIAAHWQAAAHWQQQ 357

Qy     463 ----------  ---------              MALPV    PTPNPPGYGAPVP 480
                                              |:|||     :  : :   :
Db     358 QQQPVGPQYQGIMQPHTQLQQQIPQQYPQVTQHSPPHIQPMGLPVQLPHYPVQLQLPLHIP 417

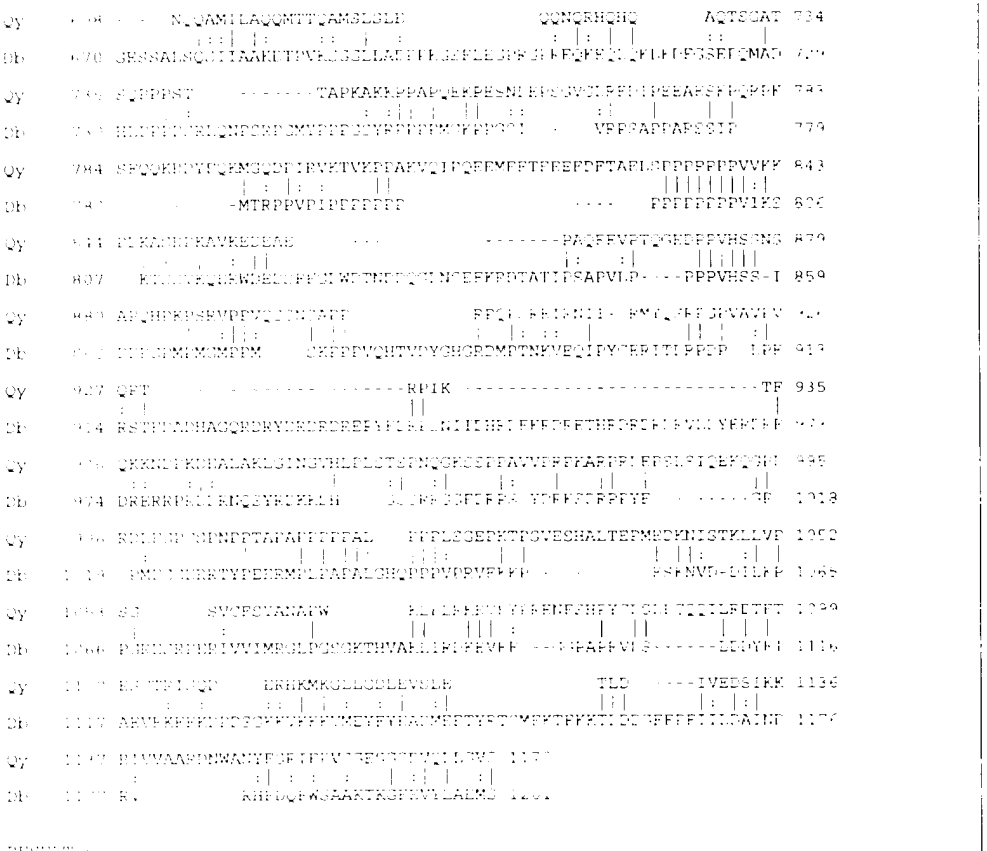
Qy     481 SWTTGPLPAGWILQGPGLEAPPPGWGVSLHG        ---GPAWPGLPG-----GDPVL 526
           |    |                  |  |||              |:  |       ::
Db     418 DWPPPPGL   ------------WPPQLHPYPHPPQLQPYPPGWPTWQGHMPATQTVIQ 461

Qy     526 DLIG---- QTPDLGDPAGPHNY---                 -------    PITPLGIAPSIPP 556
           : :       ::: ||: | | :                            |: |  |  ||||
Db     462 PPVHSPQTVPSQYLGNMAMPPPPVPYSQMPPPIPTMPPPVLPPGILPPPVMPPALPTSIPP 521

Qy     557 APGVQAPSLPPGL-----PPGPAPILASSPPP-------GEAGKPPNLGGPVPHLPPPAL 604
           ||:  | :|| |      |||  | |:|  ||           |  |           | |
Db     522 -PGMPPPVMPPSLPTSVPPPGMPPSLSSGPPPVLPPPALSSAGSP----------- PVL 568

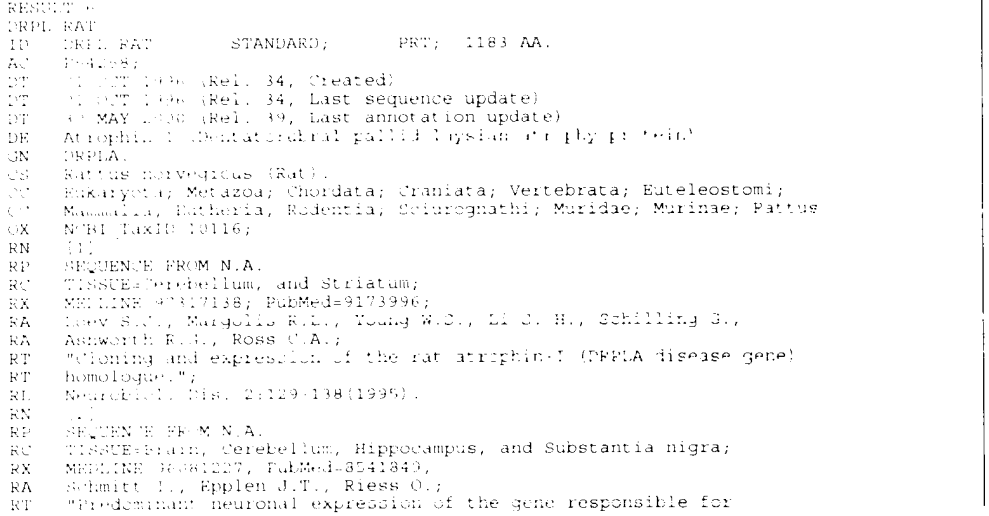
Qy     605 APGPSPDLPQGWALSPPMKGGGGSVGP--------TQQGYPMVYPPQMVQAPGYQPAMIPAPM 656
           |          : ||   |              | |   ||   ||   |  | : ||:
Db     569 PP----------PALPGGPPILPLPPLSSATPPPGIPP--PG---APQGMPPQITAPL 611

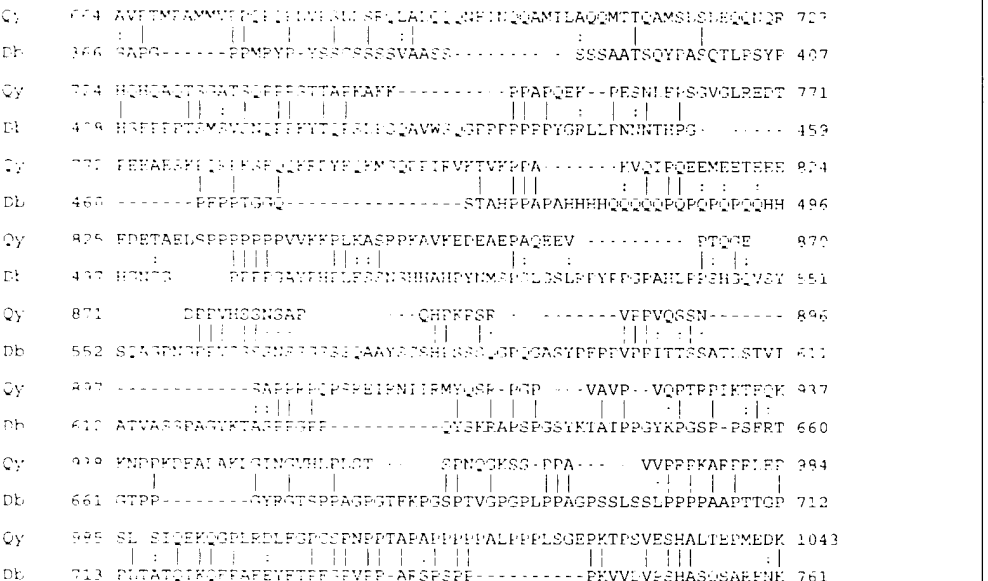
Qy     657 PVMPAMGA   -VPTMPAMMVPP--------QPQPLVPALDSPQLALQGQNFI------   697
           |  || |:    :|  |   : |          | |   |  | |:  :  | |:
Db     612 P  PASGSQNSQIPPPPQALLPTPVSPSTTPSPYHPPPQPPQIRIPPPLMPVPSSEQHL 669
```



```
RESULT 6
DRPL_RAT
ID   DRPL_RAT       STANDARD;      PRT;  1183 AA.
AC   P54358;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAY-2000 (Rel. 39, Last annotation update)
DE   Atrophin-1 (Dentatorubral pallidoluysian atrophy protein
GN   DRPLA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum, and Striatum;
RX   MEDLINE=97317138; PubMed=9173996;
RA   Lorey S.L., Margolis R.L., Young W.S., Li J. H., Schilling G.,
RA   Ainsworth R.I., Ross C.A.;
RT   "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT   homologue.";
RL   Neurobiol. Dis. 2:129-138(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX   MEDLINE=98061227; PubMed=8541849;
RA   Schmitt I., Epplen J.T., Riess O.;
RT   "'Pan-neuronal' expression of the gene responsible for
RT   dentatorubral pallidoluysian atrophy (DRPLA is ubiquitous.";
RL   Hum. Mol. Genet. 4:1619-1624(1995).
CC   -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC       MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC   -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC       (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licensing agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U31777; AAA80337.1; -.
DR   EMBL; X83548; CAA58623.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   Pfam; PF03154; Atrophin-1; 2.
DR   PRINTS; PR01222; ATROPHIN.
FT   DOMAIN       165    171       POLY PRO.
FT   DOMAIN       303    306       POLY PRO.
FT   DOMAIN       [illegible]      POLY GLY.
FT   DOMAIN       387    391       POLY SER.
FT   DOMAIN       440    446       POLY PRO.
FT   DOMAIN       477    480       POLY HIS.
FT   DOMAIN       481    489       POLY GLN.
FT   DOMAIN       502    505       POLY PRO.
FT   DOMAIN       662    672       POLY SER.
FT   DOMAIN       702    705       POLY PRO.
FT   CONFLICT     455    455       N -> S (IN REF. 1).
FT   CONFLICT     594    594       F -> L (IN REF. 2).
FT   CONFLICT     689    689       P -> R (IN REF. 2).
FT   CONFLICT     717    717       T -> M (IN REF. 2).
FT   CONFLICT     737    737       A -> V (IN REF. 2).
FT   CONFLICT     965    965       MISSING (IN REF. 2).
SQ   SEQUENCE   1183 AA;  124778 MW;  7FB9928D7A1F9B1F CRC64;

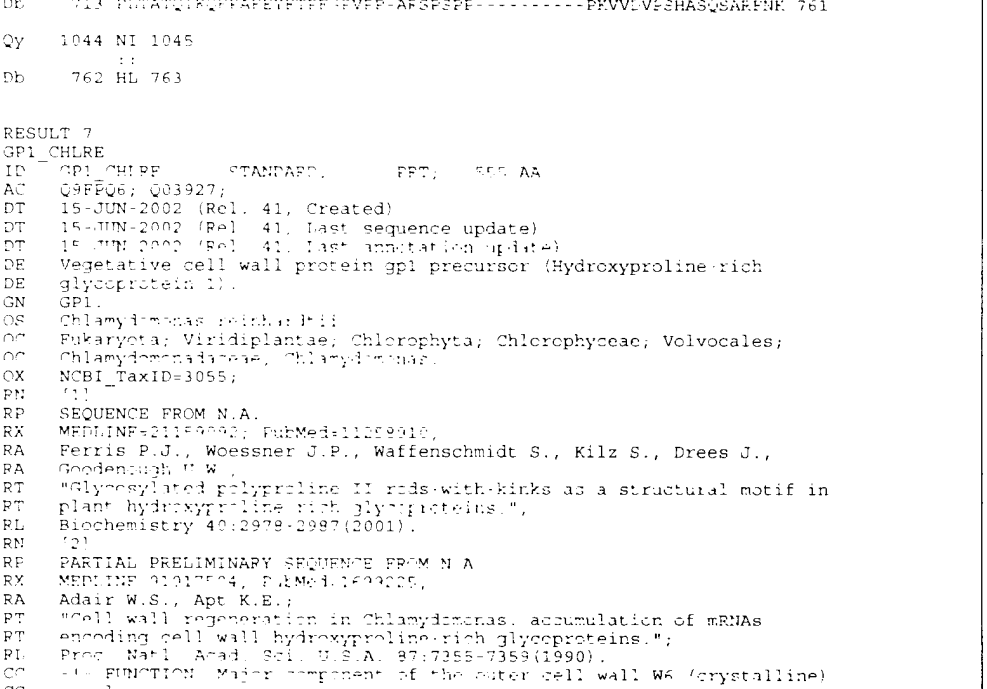
  Query Match          4.1%;  Score 206.5;  DB 1;  Length 1183;
  Best Local Similarity   21.8%;  Pred. No. 3.1e-07;
  Matches  210;  Conservative   73;  Mismatches  422;  Indels  457;  Gaps  47;
```





RESULT 7
GP1_CHLRE
ID   GP1_CHLRE   STANDARD;   PRT;   555 AA.
AC   Q9FPQ6; Q03927;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Vegetative cell wall protein gp1 precursor (Hydroxyproline-rich
DE   glycoprotein 1).
GN   GP1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21159993; PubMed=11259910;
RA   Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA   Goodenough U.W.;
RT   "Glycosylated polyproline II rods with kinks as a structural motif in
RT   plant hydroxyproline-rich glycoproteins.";
RL   Biochemistry 40:2978-2987(2001).
RN   [2]
RP   PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX   MEDLINE=91017504; PubMed=1689035;
RA   Adair W.S., Apt K.E.;
RT   "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT   encoding cell wall hydroxyproline-rich glycoproteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC   -!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC   layer.
CC   -!- SUBUNIT: Associates with GP2 and GP3.
CC   -!- PTM: N-glycosylated and O-glycosylated.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AF309494; AAL45420.1; -.
DR   EMBL; M58496; AAA69706.1; ALT_SEQ.
DR   GlycoSuiteDB; Q9FPQ6; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS01218; PSTLEXTENSIN.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL   1   29   POTENTIAL.
FT   CHAIN   30   555   VEGETATIVE CELL WALL PROTEIN GP1.
FT   DOMAIN   40   339   49 X 5 AA APPROXIMATE PPSPX REPEATS.
FT   DOMAIN   259   279   POLY-PRO.
FT   CARBOHYD   349   349   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   455   455   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   493   493   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   555 AA; 54219 MW; 6A594A9046560295 CRC64;

  Query Match   4.1%; Score 289.5; DB 1; Length 555;
  Best Local Similarity   23.2%; Prod. No. 2.8e-06;
  Matches 124; Conservative   29, Mismatches 206; Indels 175; Gaps   19;



RESULT 8
GSR1_HUMAN
ID   GSR1_HUMAN   STANDARD;   PRT; 1509 AA.
AC   Q9NZM4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glioma tumor suppressor candidate region gene 1 protein.

```
GN   GLTSCR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE 20[illegible]9456, PubMed 10[illegible]617,
RA   Smith J.S., Tachibana I., Pohl U., Lee H K., Thanarajasingam U.,
RA   Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H W.,
RA   Scheithauer B.W., Louis D N., Jenkins R B.;
RT   "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT   region.";
RL   Genomics 64:44-50(2000).
CC   -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC       placenta, skeletal muscle, and pancreas, and at lower levels in
CC       lung, liver, and kidney.
CC   ---------- ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------------------------------------------------------
DR   EMBL; AF182077; AAF62874.1; -.
DR   PDB; 1GW5; 08-NOV-01.
DR   MIM; 605690; -.
FT   DOMAIN       37     45       POLY GLY.
FT   DOMAIN      884    889       POLY-PRO.
FT   DOMAIN     1214   1225       POLY SER.
FT   DOMAIN     1282   1286       POLY-PRO.
FT   DOMAIN     1294   1304       POLY-PRO.
SQ   SEQUENCE   1509 AA,  [illegible] MW,  7251[illegible]F4483F6801 CRC64,

  Query Match             4.1%;  Score 287.5;  DB 1;  Length 1509;
  Best Local Similarity   22.2%;  Prod No  1e-05;
  Matches  214;  Conservative   75;  Mismatches  355;  Indels  318;  Gaps   42;

Qy     221 A[illegible]SFWTFVTNMFLGFLFIPA[illegible]LATLLFFAFSHQALLTGSITFSLFFFVFAFFSL 281
           ||:  ||:|       :    | :||                                |    |  :  :|:
Db     264 APTGTPSGQPLAVAPCLCGSPL-VPA-          -------PNVILHRTPTPIQPKPAG 306

Qy     282 TLPPDIPQPPPSGFVSTSPQPFFLPF       PGQPLPPPLTFLDGENPQQALEINPVML 335
           |||: :|              |||:|          |:|        |  |||       :
Db     307 VLFFKLYQLT          FRFFAFAGATLTIQSFP-         -SALPQQFFAPQNLTF 348

Qy     336 SLLGLLLLWLELTNSIPFVLLALFFLLFFSQLVAQLWFPQSPQQNDFGWALMVI 390
           |:           |  :|    | |:  :| |  |             |
Db     349 MAAGKA          QQNVVLSGPFAPALLANVFRLPATTPGAAPPQ---------- 387

Qy     391 LLSSPAITFALERPLLRPVCDLAFSDMAALCQHPILDGALPQTPLAPMASPSHFPTQLEWK 450
           |  || ||:   : :|  | :    || | |  :   | | |     | ||
Db     388       PPGALCKFNGVHLLRQG SSTVIPALHMIPP--QNQPLLPGAPAVQLPQCISAL 438

Qy     451 ALARRGKMALEVFTFREECYSAEVESWTTGEQFAGWLQGFGLFAP-FPIWFVSLHSGPA 509
           |:|:          |      | |:  |  || :: |  |    |  :: |||
Db     439 PANVCQQTL----------  AAAAFHTGQLIANFILTNQNLAGPLSLGPVLAPHSGAH 486

Qy     510 WRPLPGCDPVLLLIGG-          TEELGPPAGPHNYPI      ------- 545
           :    :   :||                |:    ||||   :
Db     487 SAHILSAAPI  QVQQPALPQMPVSLAAGSIPTQSQPAPAGPAATTVLQGVTLPPSAVAM 544

Qy     546   TPLGLAESIPPA-----------PGVQAP-----SLPPGL--------PPGPAPIL 577
             || ||  :   ||          |  |||        || ||        |  | |
Db     545 LNTPDGLVQPAIPAAATGEAAPVLTVQPAPQAPPAVSTPLPLSGAQPQAGPPQAPTPQA 604

Qy     578 AS       DEPPGEASKPENLDGFVPHLFFAIAPGFSMLFLGWALSRFMPGGGEVGP- 629
           |:          || || :| :      | :|  |   : | |  :|   :    | |
Db     605 AAPPQATTPQPSPGLASSPPI-----VLPQPPSATPTATLTQL-SLQMFLPQERSQQPL 658

Qy     630 TQQQYDMVTFGMVQAPGY[illegible]AMIIAF MEVMPAMGAVPTMPAMMVPPQPQPGLVPGLGPPQ 688
```

```
           : :|  :  |  |   :  ||  |||  ||    |  |  |     | ' | ::
Db     659 SAEGPHLSVPASVIVSAPPPAQDPAPATPVAEGAGLGPQAPIGLQASPAPAPQIPA      714

Qy     689 [illegible]NFIN[illegible]AMIIA[illegible]MIPAMGL[illegible]PGLNGRRQRQAQTGATTQFIPGTTAPKAEP 748
                     |  |     : : ||          |||    :   |  |  | | ::
Db     714 ------------AAPLKGPGPSSSPSL -----   PHQAPLGLSTHLPSPHPTKPPGPP 756

Qy     749 PPAPLEPPEPNLEIGSVSHEIIPEVAPGPPGATP-PLLPETTLLPM-LPIIEMPTVPPM 808
           |  ||          ||          |       || | :          ::|     : |:
Db     753 PSRPQSVSRPPSEP          PLHPGPPPQAPPTPPGIPVIQNQIG          VPPP 799

Qy     809 AEVQIPQEEMEETEEEEIETAELLIPPPPPEVVPEIPAAPPPGP-PPAPPAPPVPTQ 868
           |                 | :| || ||:          |:::   :    |    :|
Db     736 A------- --        ONPAPTAPGPTQPPL         NPGQQTIPNILITAIHLPPS 881

Qy     869 SEIPIVHSLNEAPLH   PPL     PVPTVAL M ALIPL P PPPNIPEMYQVPPPGP 921
           | ||:          | ||     : || |  : : | | |:  :              '|
Db     832 SISSAVASSSETSSRLPAPTPSDFQLQPPPSQGPHKS PTPPPTLHL      --   VPEP 881

Qy     922 VAVIN[illegible]PTPIPIPLPPNIPPPALAPLIPNVPPIPTTPNL PPPPPAVPP          979
           | |  | |   |||                          : |:|       | |: |
Db     882 AAPPPPPPP - TPQ                         MVTTPPPAIPQPPAIPPPPQVPP 917

Qy     980           PPAPPPLPPPLPIPPPGPPIPP       PPPPPPPPPPTAPAPPPPPA   1037
                     | |:   ||       |||         : :      :|||:  : ||
Db     918 GIILQNKAQGAPAAPQIsisl------GPLISPAASVIVQQLAIQGTTAPSHAPAPAPM 971

Qy    1038     LPPPLSGEPP        TPSVPGHAITPPMPPPNI  TPPI          VPPPPV 1094
               |||  |  | |       | :|   |  :  |  :   :  :|:        | ||
Db     972 AATGLPIILPAEGPAPA NLI TIQVAPAA-G-GPPI  G IP PL  LPPIIL  PPA 1021

Qy    1095 CF 1058
           ||
Db    1032 CF 1033
```

```
RESULT 9
DPPL_HUMAN
ID   DPPL_HUMAN     STANDARD;      PRT;  1185 AA.
AC   P54259;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Atrophin-1 (Dentatorubral pallidoluysian atrophy protein).
GN   DRPLA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum, and Brain;
RX   MEDLINE 95144175; PubMed=7842016;
RA   Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA   Inoue T., Yamada M.;
RT   "Structure and expression of the gene responsible for the triplet
RT   repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";
RL   Nat. Genet. 8:177-182(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 96[illegible]14; PubMed 8665542;
RA   Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA   Kidwai A.S., Ashworth R.G., Ross C.A.;
RT   "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT   brain.";
RL   Brain Res. Mol. Brain Res. 36:219-226(1996).
RN   [3]
RP   SEQUENCE OF 470-725 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE 98315145; PubMed=9625628;
RA   Li S. H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT   "Novel triplet repeat containing genes in human brain: cloning,
RT   expression, and length polymorphisms.";
```

```
RL   Genomics 16:572-579(1993).
CC   -!- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC       OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE
CC       LIVER, THYMUS AND LEUKOCYTES.
CC   -!- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC       (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC       ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC       EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC       DISEASE.
CC   -!- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC       PALLIDOLUYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC       DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC       NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUYS' BODY. CLINICAL FEATURES
CC       ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC       THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC       IN THE FOURTH.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collabora-
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D31840; BAA06276.1; -.
DR   EMBL; U23851; AAB50276.1; -.
DR   EMBL; L10377; -; NOT_ANNOTATED_CDS.
DR   Genew; HGNC:3033; DRPLA.
DR   MIM; 125370; -.
DR   InterPro; IPR002951; Atrophin.
DR   Pfam; PF03154; Atrophin-1; 2.
DR   PRINTS; PR01222; ATROPHIN.
KW   Triplet repeat expansion; Polymorphism.
FT   DOMAIN       73     82       SER/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      302    305       POLY-PRO.
FT   DOMAIN      376    382       POLY-SER.
FT   DOMAIN      386    397       POLY-GLY.
FT   DOMAIN      442    447       POLY-PRO.
FT   DOMAIN      479    483       POLY-HIS.
FT   DOMAIN      484    497       POLY-GLN.
FT   DOMAIN      504    507       POLY-PRO.
FT   DOMAIN      564    574       POLY-SER.
FT   DOMAIN      594    597       POLY-PRO.
FT   DOMAIN      802    815       ARG/ALA-RICH (MIXED CHARGE).
FT   DOMAIN      816    827       ARG/GLY-RICH (MIXED CHARGE).
FT   DOMAIN      925    934       ARG/GLU-RICH (MIXED CHARGE).
FT   CONFLICT     94     94       MISSING (IN REF. 2).
FT   CONFLICT    333    333       Y -> H (IN REF. 2).
FT   CONFLICT    339    339       M -> I (IN REF. 2).
FT   CONFLICT    541    541       P -> T (IN REF. 3).
FT   CONFLICT   1028   1028       G -> A (IN REF. 2).
SQ   SEQUENCE   1185 AA;  124[illegible] MW;  5[illegible] CRC64;

  Query Match             4.1%;  Score 286.5;  DB 1;  Length 1185;
  Best Local Similarity   22.5%;  Pred. No. 8.6e-06;
  Matches  216;  Conservative   91;  Mismatches  313;  Indels  349;  Gaps   52;

Qy    223 APGPHS       GPPWGPVPNMDLGPLPIPAQLA  TLLPPAPGHQALLTG   SITES 278
Db     20 APGPPPELPPPGPASPGGVSTSSSPPKPAEKSPQTAPPAPVPPASTPPVNPQGPPPPISPS 79

Qy    279 IPPPVPAPPSLTLPPSIPPPSPVSTSPLPPPLPPPQPPLSPPLTPLPS ------ P 322
Db     80 ESEETNA-PKKTK----------------TPQPLPPPQSPSD--LDSLDGPSLNDDGSS 119

Qy    323 NPSQAIPIHP VMLPIPPPTPMPWPGLIMSTPLVQLAQPPPGLPHPLPSQLVAQLWPNP 381
Db    120 DPRDIDQDNRSISPSIYSPGPVPNDSDSSG---LSQGPAPPYHPPPLPPPSP-----QPP 172

Qy    382 PPQLNGPGWAIMVILISPAITPA        LEPILLKPVSDQAPSGMAALCQHPLL- 431
Db    173 DGTPPGPP        ASPPPHPPVTPTTGYHAPMPPTTSPMP -QAPPG--AIPHPQGY ..
```

```
Qy    432 -----GALPQTPLAPMASPSHPPTQLEWPASLPPGPMALPVPTPNPLPTSAATVLWTTGN 486
Db    321 PGSPPGVLPGPPMGP          -         - -PDGGAASSVGGPN                 339

Qy    487 QPAGWILQSPGLEAPPPGWGVSLHSGDAWPLPGGPPVLPLLPGTPPLGPPAGPHNYPIT 546
Db    340 -         - DGPQHPPPTTPISVSSSGA- -                        SGAPTTKPPTT 378

Qy    547 PLP   LAPSPPAPGVQAPLPLPPSPA  PI LAPPPPIPAPPPNLLPTPHLP 600
Db    379 PVSGGNLPGAPPPA -NPTPNTTNLPPTPALPTNNAGAPPPGLGAQP--LPG-- HLP 424

Qy    601 PPALAPGPSPLEQGWALSPRMPGGSVGHTQQGTPMVTPMVLAPLTPLANIPAPMPTMP 660
Db    425 SP------------YAMGQGMGG-------------LPPG----PPAGPILAPSPHSLPP 463

Qy    661 AMGAVPTMTAMMVPIPQPIVPSLGGPLPALPLNPPIPLAMIPAPLMIPPANGLGLPQQ 720
Db    464 ASPSAP-   ---APPMRPPYSSSSSSSAAASSSSSSSSSA---SPPPASQAL-     P 506

Qy    721 MAPHPAPAPSGATSGPPPPGTTAP        -PAPPPPAPPPP  PPPNLPPGGVGLRP 769
Db    507 SYPHSPPPPPTGLSVGNPPPPYTPPLPSPAVWSPPPPPPPYGPLLANGNAHPG -     463

Qy    770 PTPPPAPGPPQPPPSPPQPPPDYPPMGPPPSVPTVPPPAPV---QIPPEEMPPTPPPPP 826
Db    462 ---------PPPPSTGAQ----------------STAPPPVSTHHHHHQQQQQQQQQQQQ 496

Qy    827 ETAELSPPPPP----------------PPVVKKPLKASPPPAVKEDEAEPAQEEV---PT 867
Db    497 QHHRNPPPPPPSAPPHPPPPGSSPHHHAHPVAMSPSPGPSLSPYPPPGPAHPPPPPPVSYSQA 556

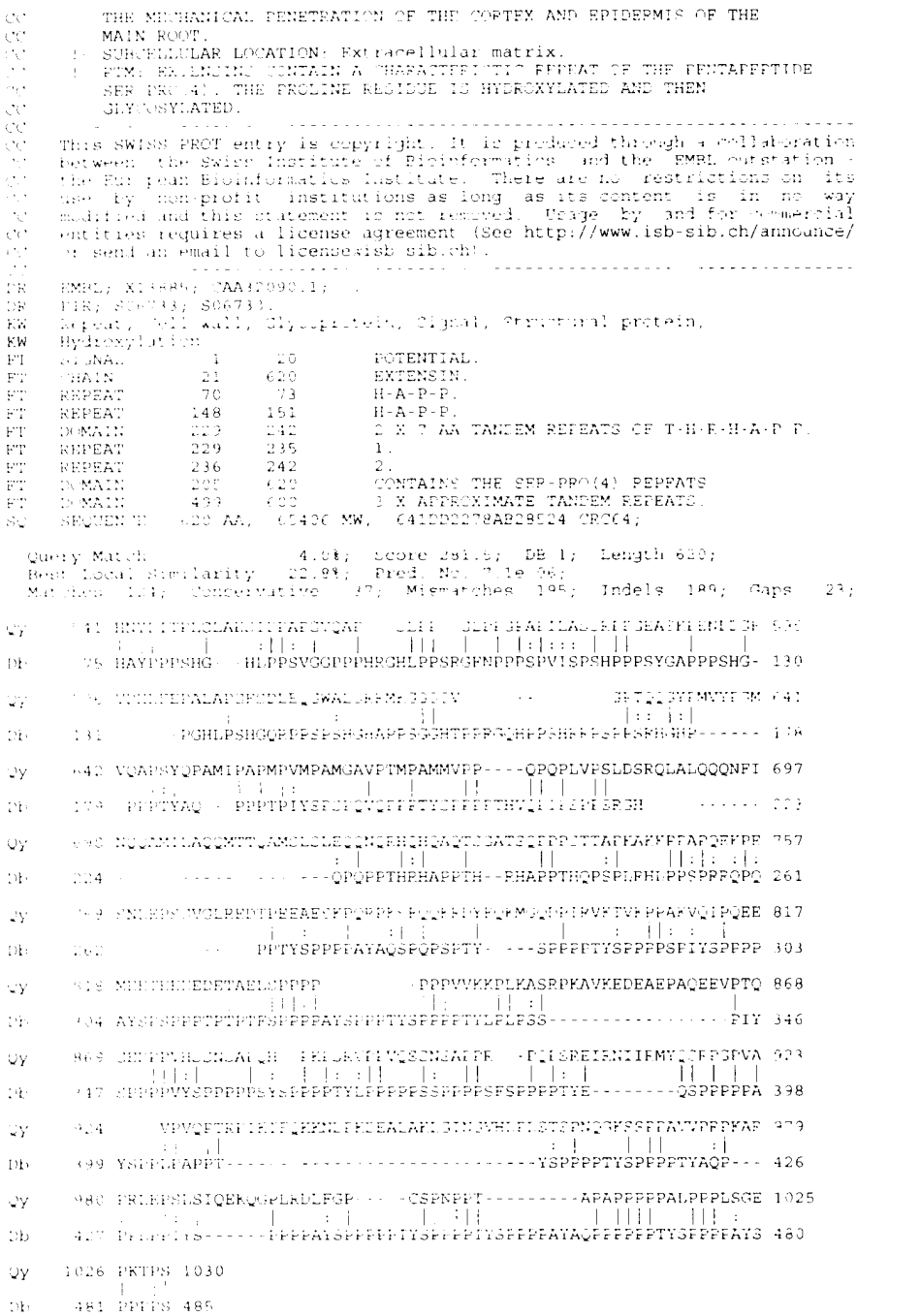
Qy    868 QGEPPPVHGSNGAP           QHPKPSR--  - -- --YPPVQGGN------      896
Db    557 GPNGPPVSPGSNNSSSGTSQGSYPGPHPSPGPPGPGAPVPPPPPVPTVTTGGATLGTTATV 616

Qy    897  -------SAPPPPQPSPEIPNIIPMVPGP PGP     VAVP  VIPTPPPPTPGPPNP 940
Db    617 APSPAVYPTASPPGPPP----------YGPPAPPPGAVPTATPPAPPPPSPPPPGTP 665

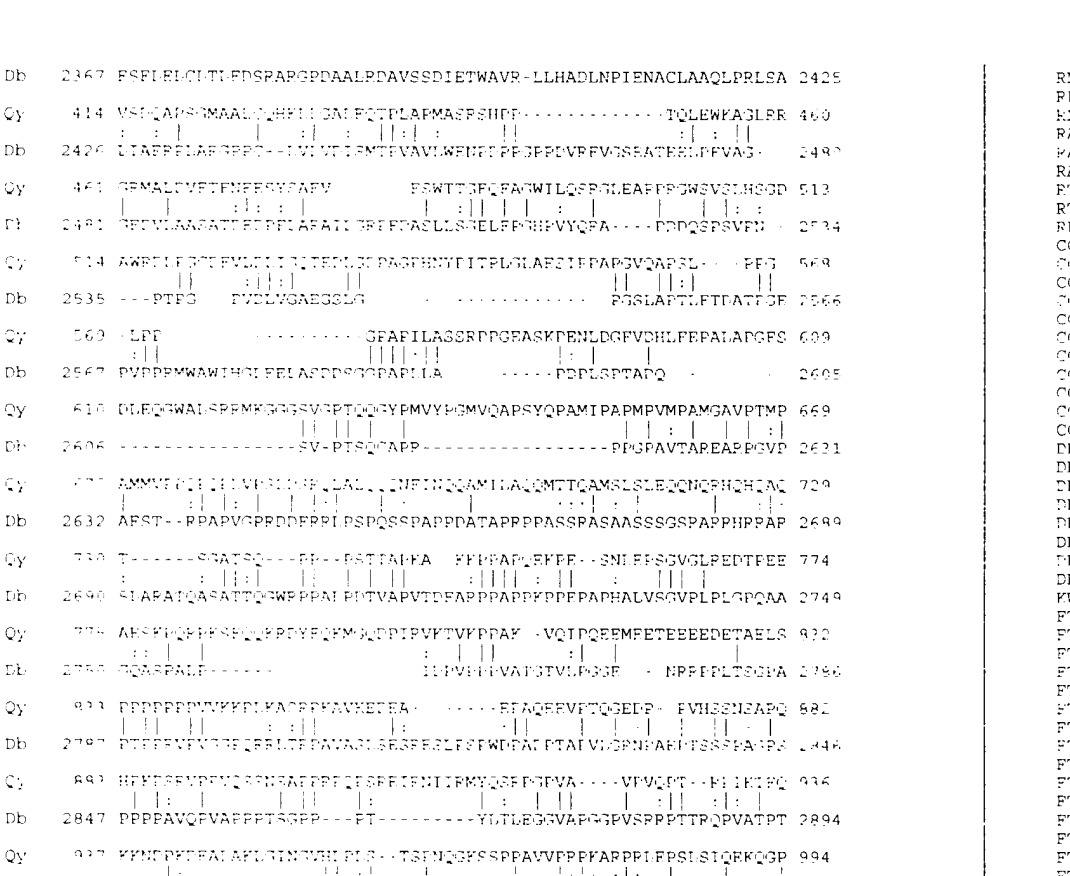
Qy    941 PPPEALAPLPGINGVHLPLPT---  -PPNQHGG PPA ---- VVPPPPAPPPLPPPGLS 987
Db    666 P------- -GYPGTPPPAGPGTPPPPGPTVGPGPLPPAGPGGLPGLPPPPAAPASGPPPLS 717

Qy    988 -IPPPQGPLPGLPPPGPPPPPPAPAPPPPPPALPPPLSGPPPTPPAVPLHALTPPMPPPP 1046
Db    718 ATQIPQGPPAPPYPTPLSPVPP-ARGPSPP--- ---  --PPVVPVPPHAPQGARPPNKHL 765
```

```
RESULT 10
EXTN_TOBAC
ID   EXTN_TOBAC     STANDARD;      PRT;   620 AA.
AC   P13983;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
GN   HRGPNT3.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN cv. Xanthi, TISSUE=Leaf;
RX   MEDLINE=92128763; PubMed=2612909;
RA   Keller B., Lamb C.J.;
RT   "Specific expression of a novel cell wall hydroxyproline-rich
RT   glycoprotein gene in lateral root initiation.";
RL   Genes Dev. 3:1639-1646(1989).
CC   -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
```

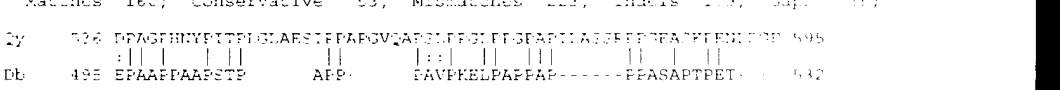

THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
MAIN ROOT.
-!- SUBCELLULAR LOCATION: Extracellular matrix.
-!- PTM: EXTENSINS CONTAIN A TRANSLATIONAL REPEAT OF THE PENTAPEPTIDE
SER PRO 4, THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
GLYCOSYLATED.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X14885; CAA32911.1; -.
PIR; S06733; S06733.
Repeat, Cell wall, Glycoprotein, Signal, Structural protein.
Hydroxylation.
POTENTIAL.
EXTENSIN.
H-A-P-P.
H-A-P-P.
2 X 7 AA TANDEM REPEATS OF T-H-E-H-A-P-P.
1.
2.
CONTAINS THE SER-PRO(4) PEPPATS
3 X APPROXIMATE TANDEM REPEATS.
SEQUENCE 620 AA; 65406 MW; 6410DD278AB28524 CRC64;
Query Match 4.0%; Score 291.5; DB 1; Length 620;
Best Local Similarity 22.0%; Pred. No. 7.1e-56;
Matches 121; Conservative 47; Mismatches 195; Indels 189; Gaps 23;

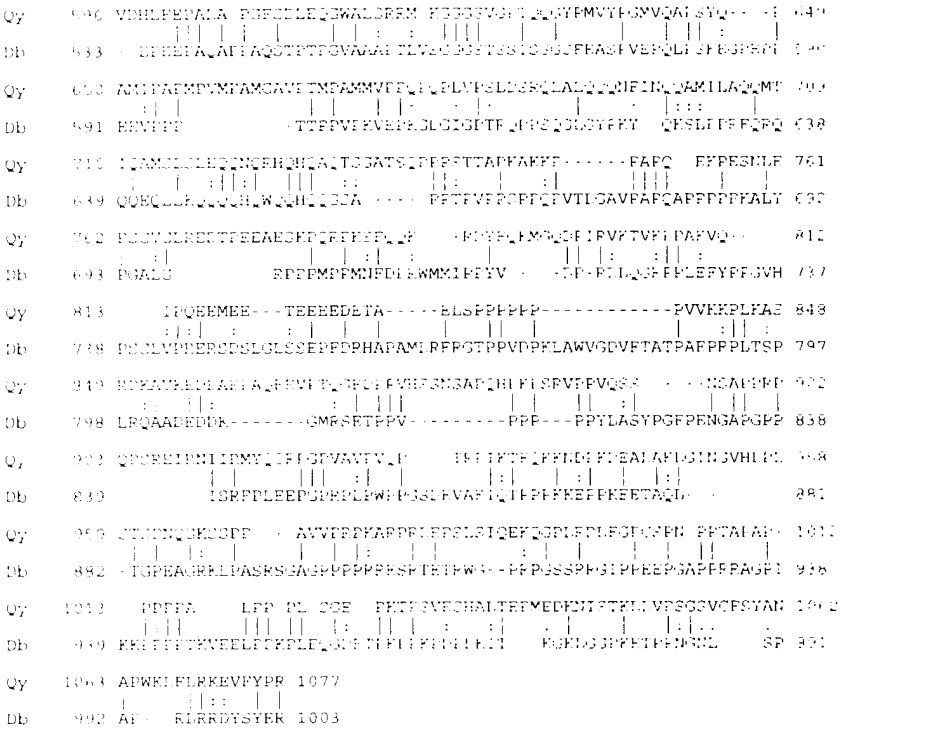
RESULT 11
TEGU_HSV11
ID TEGU_HSV11 STANDARD; PRT; 3164 AA.
AC P10220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Virion protein VL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses, dsDNA viruses, no RNA stage, Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; X14112; CAA32311.1; -.
DR PIR; I30085; WMREH6.
DR InterPro; IPR005210; Herpes_UL36.
DR Pfam; PF03586; Herpes_UL36; 1.
KW Repeat.
FT DOMAIN 2911 2986 35 X 2 AA TANDEM REPEATS OF T-S-.
SQ SEQUENCE 3164 AA; 335857 MW; CC5D31FF4AF9FE4F4 CRC64;
Query Match 4.0%; Score 280.5; DB 1; Length 3164;
Best Local Similarity 21.5%; Pred. No. 4.4e-55;
Matches 241; Conservative 104; Mismatches 419; Indels 402; Gaps 17;



RESULT 12
BAT2_HUMAN
ID   BAT2_HUMAN   STANDARD;   PRT; 2142 AA.
AC   P48634;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Large proline-rich protein BAT2 (HLA-B-associated transcript 2).
GN   BAT2.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=90192810; PubMed=2156268;
RA   Banerji J., Sands J., Strominger J.L., Spies T.;
RT   "A gene pair from the human major histocompatibility complex encodes
RT   large proline-rich proteins with multiple repeated motifs and a
RT   single ubiquitin-like domain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
RN   [2]
RP   SEQUENCE OF 1-1860 FROM N.A.
RX   MEDLINE=93172029; PubMed=8499947;
RA   Iris F.J.M., Bougueleret L., Prieur S., Caterina D., Primas G.,
RA   Perret V., Jurka J., Rodrigues Tome F., Claverie J.-M., Dausset J.,
RA   Cohen D.;
RT   "Dense Alu clustering and a potential new member of the NF kappa B
RT   family within a 90 kilobase HLA class III segment.";
RL   Nat. Genet. 3:137-145(1993).
CC   -!- FUNCTION: UNKNOWN.
CC   -!- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; M33509; AAA35585.1; -.
DR   EMBL; M33518; AAA35586.1; -.
DR   EMBL; M33512; AAA35586.1; JOINED
DR   EMBL; Z15025; CAA78744.1; -.
DR   PIR; B35098; B35098.
DR   PIR; S36152; S36152.
DR   Genew; HGNC:12018; BAT2.
DR   MIM; 142580; -.
KW   Repeat.
FT   DOMAIN   519   524   POLY-PRO.
FT   DOMAIN   636   657   GLN-RICH.
FT   DOMAIN   684   688   POLY-PRO.
FT   DOMAIN   699   704   POLY-PRO.
FT   DOMAIN   814   821   POLY-PRO.
FT   DOMAIN   1340   1345   POLY-GLY.
FT   DOMAIN   1398   1403   POLY-GLY.
FT   DOMAIN   1426   1441   POLY-PRO.
FT   DOMAIN   1982   1991   POLY-PRO.
FT   DOMAIN   41   1795   4 X 57 AA TYPE A REPEATS.
FT   REPEAT   41   95   1-1.
FT   REPEAT   98   154   1-2.
FT   REPEAT   281   337   1-3.
FT   REPEAT   1740   1795   1-4.
FT   DOMAIN   337   549   2 X TYPE B REPEATS.
FT   REPEAT   337   418   2-1.
FT   REPEAT   476   549   2-2.
FT   DOMAIN   1894   2089   4 X 50 AA TYPE C REPEATS.
FT   REPEAT   1899   1948   3-1.
FT   REPEAT   1965   2014   3-2.
FT   REPEAT   2040   2089   3-3.
FT   CONFLICT   57   57   P -> A (IN REF. 2).
FT   CONFLICT   109   109   Q -> S (IN REF. 2).
FT   CONFLICT   414   414   P -> PPHRGPAGNWGPP (IN REF. 2).
FT   CONFLICT   532   532   T -> P (IN REF. 2).
FT   CONFLICT   602   602   Q -> K (IN REF. 2).
FT   CONFLICT   730   730   E -> D (IN REF. 2).
FT   CONFLICT   750   750   L -> R (IN REF. 2).
FT   CONFLICT   834   834   A -> T (IN REF. 2).
FT   CONFLICT   1035   1035   G -> A (IN REF. 2).
FT   CONFLICT   1068   1068   M -> L (IN REF. 2).
FT   CONFLICT   1285   1285   P -> R (IN REF. 2).
FT   CONFLICT   1400   1400   G -> A (IN REF. 2).
FT   CONFLICT   1611   1611   T -> S (IN REF. 2).
FT   CONFLICT   1729   1729   G -> A (IN REF. 2).
SQ   SEQUENCE   2142 AA; 227840 MW; [illegible] CRC64;

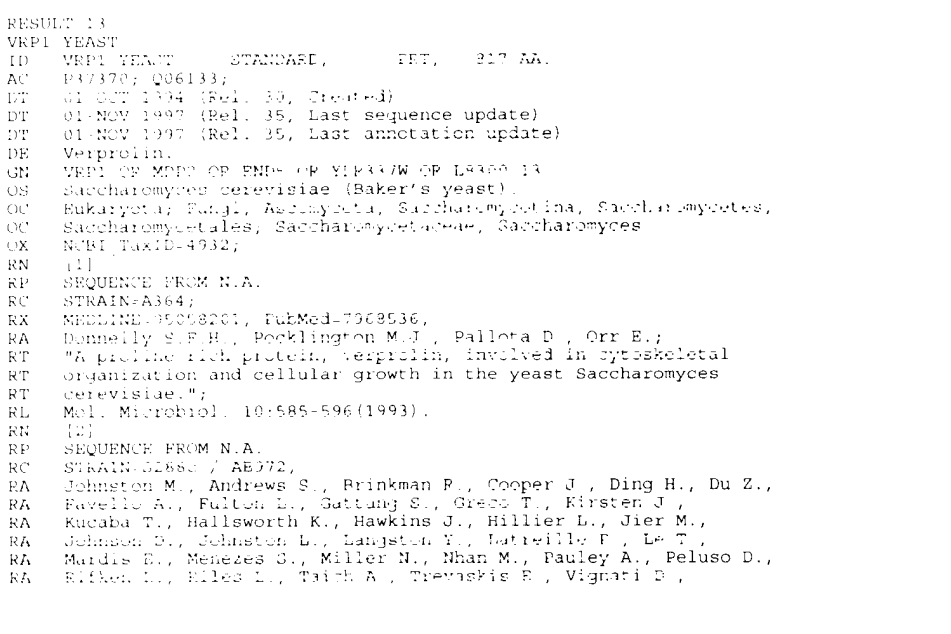
  Query Match   4.0%; Score 276.5; DB 1; Length 2142;
  Best Local Similarity   26.0%; Pred. No. 4.4e-05;
  Matches 160; Conservative   53; Mismatches 223; Indels 173; Gaps   29;

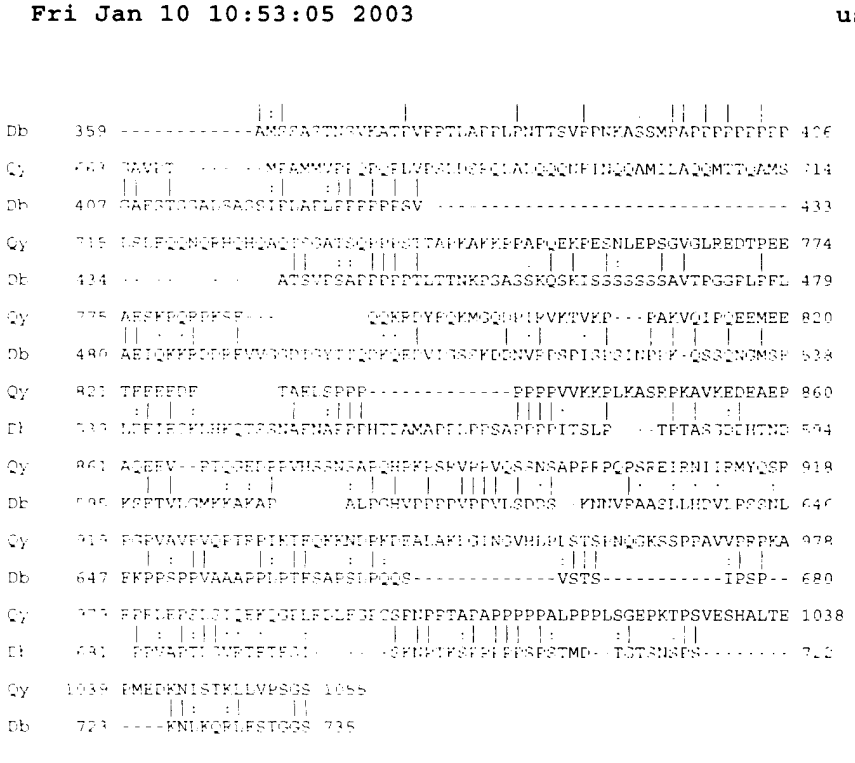




RESULT 13
VRP1_YEAST
ID   VRP1_YEAST   STANDARD;   PRT;   817 AA.
AC   P47370; Q06133;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Verprolin.
GN   VRP1 OR MOP2 OR END6 OR YIR33/W OR L9360.13
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A364;
RX   MEDLINE=95058201; PubMed=7968536;
RA   Donnelly S.F.H., Pocklington M.J., Pallotta D., Orr E.;
RT   "A proline rich protein, verprolin, involved in cytoskeletal
RT   organization and cellular growth in the yeast Saccharomyces
RT   cerevisiae.";
RL   Mol. Microbiol. 10:585-596(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifkin L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (JAN-1995) to EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN CYTOSKELETAL ORGANIZATION AND CELLULAR
CC   GROWTH. MAY EXERT ITS EFFECTS IN THE CYTOPLASM DIRECTLY, OR
CC   INDIRECTLY VIA PROLINE-BINDING PROTEINS (E.G. PROFILIN) OR
CC   PROTEINS POSSESSING SH3 DOMAINS.
CC   ---------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------
DR   EMBL; D78648; BAA01388.1; -.
DR   EMBL; U19028; AAB67263.1; -.
DR   PIR; S38626; S38626.
DR   SGD; S0004329; VRP1.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 2.
DR   SMART; SM00246; WH2; 2.
KW   Cytoskeleton; Repeat.
FT   DOMAIN   5   14   POLY-PRO.
FT   DOMAIN   239   245   POLY-PRO.
FT   DOMAIN   389   397   POLY-PRO.
FT   DOMAIN   396   406   POLY-PRO.
FT   DOMAIN   424   431   POLY-PRO.
FT   DOMAIN   462   468   POLY-SER.
FT   DOMAIN   774   788   POLY-PRO.
FT   CONFLICT   308   308   P -> R (IN REF. 2).
FT   CONFLICT   350   350   A -> R (IN REF. 1).
FT   CONFLICT   683   683   V -> E (IN REF. 1).
FT   CONFLICT   [illegible]   [illegible]
FT   DVGRYTIGGDNSIVGAKGYNERIVIIGNRFFKWCNVVQMIKI
FT   [illegible]
FT   PVKTLNNGYFLQVDRFCNTSIIRIQINQMLM (IN REF.
FT   1).
SQ   SEQUENCE   817 AA; 82593 MW; 2407522D5B1CA178 CRC64;

  Query Match   3.9%; Score 272.5; DB 1; Length 817;
  Best Local Similarity   20.8%; Pred. No. 2.4e-09;
  Matches 191; Conservative   76; Mismatches 315; Indels 335; Gaps   47;





```
RESULT 14
PCAP_HUMAN
ID   PCAP_HUMAN     STANDARD;      PRT;   746 AA.
AC   Q...N5; ...413; Q96IH7; Q96CT0; Q9P1T3;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   P...ipin-2 family 2 glutamine/Q-rich-associated protein (PC2
DE   glutamine/Q-rich-associated protein) (TPA-inducible gene-1) (TIG-1)
DE   (CTG7a).
GN   PCQAP OR TIG1
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta, and Megakaryocytes;
RX   MEDLINE=...; PubMed=...;
RA   Abraham S., Solomon W.B.;
RT   "A novel glutamine rich putative transcriptional adaptor protein
RT   (TIG-1), preferentially expressed in placental and bone-marrow
RT   tissues.";
RL   Gene 255:389-400(2000).
RN   [2]
RP   SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.
RC   TISSUE=Cervical carcinoma;
RX   MEDLINE=...; PubMed=...;
RA   Forti L., Mittler G., Przemeck G.K.H., Stelzer G., Guenzler P.,
RA   Amati F., Conti F., Dallapiccola B., Hrabe' de Angelis M., Novelli G.,
RA   Meisterernst M.;
RT   "Isolation and characterization of a novel gene from the DiGeorge
RT   chromosomal region that encodes for a mediator subunit.";
RL   Genomics 74:320-332(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kikuchi A., Hirose T., Fukuhara N., Inagaki H., Ikeda Y., Okamoto S.,
RA   Okitani F., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye, and Kidney;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 185-531 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Ridwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
CC   -!- FUNCTION: May function as a transcriptional coactivator in RNA
CC       polymerase II transcription. As a protein complex component it may
CC       regulate gene transcription, especially following induction of
CC       protein kinase C activity.
CC   -!- SUBUNIT: Subunit of the large multiprotein complex PC2.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed in all tissues examined, including
CC       heart, brain, lung, spleen, thymus, pancreas, blood leukocyte and
CC       placenta. However, the level of expression varied, with highest
CC       expression in the placenta and peripheral blood and lowest in the
CC       pancreas and kidney.
CC   -!- INDUCTION: By 12-O-tetradecanoylphorbol-13-acetate (TPA).
CC   -!- POLYMORPHISM: The poly-Gln region from amino acid ... of
CC       PCQAP is polymorphic. There are from 15 to 18 repeats in the
CC       Italian population.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to
CC       frameshifts in positions 13, 558 and 707.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to sequencing
CC       errors.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF056191; AAC12944.1; ALT_FRAME.
DR   EMBL; AF228763; AAK58423.1; -.
DR   EMBL; AK074268; BAB85034.1; ALT_SEQ.
DR   EMBL; BC013985; AAH13985.1; -.
DR   EMBL; BC007529; AAH07529.1; -.
DR   EMBL; U80745; AAB91443.1; -.
DR   Genew; HGNC:14248; PCQAP.
KW   Transcription regulation, Activator, Nuclear protein, Polymorphism;
KW   Triplet repeat expansion.
FT   DOMAIN      505     522       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      161     174       POLY-GLN.
FT   DOMAIN      178     193       POLY-GLN.
FT   DOMAIN      205     218       POLY-GLN
FT   DOMAIN      226     239       POLY-GLN
FT   DOMAIN      243     260       POLY-GLN.
FT   DOMAIN      266     473       PRO-RICH.
FT   DOMAIN      358     365       POLY-GLN
FT   DOMAIN      407     414       POLY-PRO
FT   DOMAIN      560     569       POLY-PRO
FT   VARIANT     260     260       Q -> QQQ.
FT                                 /FTId=VAR_013136.
FT   CONFLICT     12      12       S -> R (IN REF. 1).
FT   CONFLICT    116     116       L -> F (IN REF. 1).
FT   CONFLICT    161     161       Q -> P (IN REF. 3).
FT   CONFLICT    195     196       QQ -> RL (IN REF. 5).
FT   CONFLICT    231     283       MISSING (IN REF. 3).
FT   CONFLICT    263     263       Q -> E (IN REF. 1 AND 5).
FT   CONFLICT    363     381       QQQTAVQTAQAAQMVAIGV -> TAVPAIPSCCIPVCPQPM
```

```
FT                                A (IN REF. 4, AAH07509).
FT   CONFLICT    520    521       IL -> GI (IN REF. 5).
FT   CONFLICT    643    643       L -> V (IN REF. 3).
SQ   SEQUENCE   746 AA;  82324 MW;  5B8B0B6B0CAF0A3C6 CRC64;

  Query Match             3.8%;  Score 268.5;  DB 1;  Length 746;
  Best Local Similarity   21.3%;  Pred. No. 3.2e-05;
  Matches  134;  Conservative   92;  Mismatches  191;  Indels  213;  Gaps   30;

Qy     647 PLPLAETIPPAPGVLALPLPLGLPFGFAPILASGFFPGFASFPENLDGFVPHLFEFALAP 826
           |:   :|:   |   |  :  |:||          | ||::
Db      91 PMPALQTLTPPPAAVAAGI- GMPP- -       PGPGQS           ---------- 199

Qy     827 GPSDLPQGWALSRPMPGGGSVGPPQQGYPMVYPGMVQAPSYQPAMIPAPMPVMPAMGAVP 986
                       : | ||:|    | ||    | | |     | |  | |:
Db     110 ----    -------LGGMGSLG--AMGQPMSLSG--QPPPGTSGMAPHSMAVVST----- 146

Qy     987 TMPAMMVPPQPQLPVPSLDSRQLAIQQQNFINQQAMILAQQMTT 1130
           |  ||      |  |:||    ||| |  ||   | ||       :|| |
Db     147 ------ATPQTQ-----LQLQQVAIQQQQQQQQFQQQQQQAAIQQQQ---------QQQQQQ 195

Qy    1131 RHQHQAITSGATSQPPPSTIAPFAPFPDAPGKPGPSFNLPPGGVKLPELTPEEAERKPGPP 1290
           : | ||. |                    |:: .. ..    :  : . :
Db     196 QQQPQAIQQ                 -AMQQLPQAVVQQ---- -QQQQQQQQQQQQHLI 231

Qy    1291 FGPQQKPDVPQPMGGHPIPVPIVPPPAPVQIPQPFEMPFTEPPEDETAPISPP-------- 1450
           |   :   |: |   |:      |::|: |:: :: :::: :  :  ||
Db     232 KLHHQNQQQQIQPQPPQLQPI -  - AIQLQQQQQQQQQQQQQQQQALQAQPFIQQPPMQQ 290

Qy    1451 PPPPPVKLPLPAGIPAVPPPGAPIAGPPVIPVHGDIPVHSGNPAPIHPPPSPVPPVQS 1610
           |  |,     |                :  |:   ||   ||             |:| : |  |:
Db     291 PQPPPSQALP--------------QQLQQMHHIQHHQPP----------PQPQQPPVAQN 341

Qy    1611 SNPAPPSPQPGSRPINNIPMTIPPPG                PVAVPVQPTPPIPTPQPPND 1770
           |    |   | : : ::   |: ||                |: |   | :| :  |::
Db     342 QPSQLP---PQSQTQPLVSQAQALPPGMLPTTQPPLKPVPAPMVVQQPPVQPQVQVQQQTA 398

Qy    1771 PPIPNIAPTPINGVH     LPLSTGPPQG------PGSPPAVVPPPP-APPPLPPSLSI 1920
           :    |::  ||     ||. .|| .|      .| ||   | |.  .|  .|. ..
Db     399 VQTAQAAQMVAPGVQVQQQPSLPMLQGPSPQPQQVQTPQMPPPPQPSPQIQQPSSQPNSNV 458

Qy    1921 QPKQSPLPLPLPFGPGSPNSPI APAPIPTIALIPPLPGSPKTPSVESHA LTEPMELPNIS 2100
                 || :|:|  :  |:| | |:  |  :  |:  || |   |  |:   ::
Db     459 SS  -  ----GP-APSPSSFLPSPSPQPSQSPVTARTPQNFSVPSPGPLNTPVNPSSV- 476

Qy    2101 TPLLVTGSTVGPSYANAFWPLPPEVPVPPENPSHPYPLSLLPQQILPDTPTESCTRIS 2160
           : |:||     : |  : :| |              | | :|  :          :|
Db     477 -  MSPAGS-----SQAEEQQYLDK-------------LKQLSKYI--EPLRPMINKID 512

Qy    2161 QDERHKMAGLLGDLEVSLETLDIVEDSIKK 2250
           ::|  |      ||      |||: |  |:
Db     513 KNEDPFF-----DLSKMKSLLDILTDPSFP 537


RESULT 16
ZAP3_HUMAN
ID   ZAP3_HUMAN     STANDARD;      PRT;  1822 AA.
AC   P49750; P49752; Q9P1V7;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear protein ZAP3 (ZAP113)
GN   ZAP3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,
RA   Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA   Kant A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
RT   "Sequencing of human chromosome 14q24.3 region.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 533-847 AND 1577-1822 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95319502; PubMed=7596406;
RA   Chelly J., Kaplan J.C., Maire P., Gautron S., Kahn A.;
RT   "Transcription of myosin light chain 1 and 3 genes is induced by
RT   denervation during muscle regeneration: sequence, mapping, and
RT   transcription start point determination.";
RL   J. Cell Biol. 107:2177-2189(1988).
RN   [3]
RP   SEQUENCE OF 1-1822 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95319502; PubMed=7596406;
RA   Chelly J., Kaplan J.C., Maire P., Gautron S., Kahn A.;
RT   "Transcription of myosin light chain 1 and 3 genes is induced by
RT   denervation during muscle regeneration: sequence, mapping, and
RT   transcription start point determination.";
RL   J. Cell Biol. 107:2177-2189(1988).
RN   [4]
RP   SEQUENCE OF 1-1822 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95319502; PubMed=7596406;
RA   Chelly J., Kaplan J.C., Maire P., Gautron S., Kahn A.;
RT   "Transcription of myosin light chain 1 and 3 genes is induced by
RT   denervation during muscle regeneration: sequence, mapping, and
RT   transcription start point determination.";
RL   J. Cell Biol. 107:2177-2189(1988).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1761.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF007556; AAH07509.1; -.
DR   EMBL; L40404; AAC42008.1; ALT_FRAME.
DR   EMBL; S68403; AAD47175.1; -.
KW   Nuclear protein.
FT   DOMAIN       15    205       PRO-RICH.
FT   DOMAIN      382    430       GLN-RICH.
FT   DOMAIN      857   1255       APT-RICH.
FT   DOMAIN     1488   1577       ARG-RICH.
FT   CONFLICT    521    521       P -> S (IN REF. 2).
FT   CONFLICT   1404   1404       T -> I (IN REF. 2).
FT   CONFLICT   1801   1801       P -> F (IN REF. 2).
SQ   SEQUENCE   1822 AA;  204947 MW;  8B0B0B2FB540D7D2 CRC64;

  Query Match             3.7%;  Score 261;  DB 1;  Length 1822;
  Best Local Similarity   21.5%;  Pred. No. 0.00018;
  Matches  130;  Conservative   61;  Mismatches  281;  Indels  190;  Gaps   24;

Qy     535 GDPAGPHNYPITPLGLAESIPPAPGVQAPSLPPGLPPGPAPILANSGFTPSKASKPENLDG 594
           |   | :||  |       :|| | | |  |    |  :|| |   |     |
Db       6 GRTGGSSHYPPPP------VPPPPPVALPEASPG- --- PSTSRNTTPAAPSS    80 49

Qy     595 PV----DHLPEPALAPGPSDLEQGWALSRPMKGGSVGPTQQGYPMVYPGMVQAPSYQPA 650
           |:    ||      : |:|           ||  :      ::|      |
Db      50 PMSFREQHL---------AQLQQ-------------  LQQMHQPQMQPVLQPHHLHPP 84

Qy     651 MIPAPMPVMPAMGAVPTMPAMMVPPQPQPLVPSLDSRQLAIQQQNFINQQAMILAQQMTT 710
           :| | ||||  |     |    || | |  |      |: |:    | :   :
Db      85 PLPPP-PVMPGGGSYGLWLP----PPPIMPPPP  GPALSTQPQQQQPKHQMLHHQRDGPP 136

Qy     711 PAMSLLELPPNL        PRLRLALSLSATSQPPPLTIAPFAPFPDAPGKPGPSFNLPPGGVKLPELTPEEAERKPGPP 790
           : :||   :          :   :|:    |||||   | : :|  |  :|  :  |
Db     137 GLVPMELESPPESPPVPPGSYMPPSQSYMPPPQPPPGYYPPTSNQIYLPPAQPSPSQSPP 196

Qy     791 SVGLPEPTPEEAEPKPLPPPSPLTLPMPLIPPPVP                TVKPPA 850
           |  : ||  :|  :  :  | : :| | |  :                    :  |
Db     197 SQSYLAPTPSYSSSSSSQSYLSHSQSYLPPSQASPSRPSQQHSKEQLLAPPPPQSAPPQN 256

Qy     851 FVQIPLPPPMPE-IPPPPPPIAPLSPPPPPPPV- ---            VPPPLPKAGKP 907
           |  : ||  :|  :  :  |  : :| | |  :                 :::  |
Db     257 PTTVLPPPLPLPAPNPSTPLLLAAPPPPPPPQMIPLPLLLWPPLPPLPLPTPVHLPQH 316

Qy     908     FAVPEIEAPPAQPPVPIPGPLPLGPSLNLAIPPPPPPPPPPPPPPAPLPPQLPPE 967
               | :|  || :|||                  |  |:||   ||   || | :  |
Db     317 PPGPVVAPDTPPPVPPPVTV               PATPQVPPESPSSEP PPLPPPNEE 369

Qy     968 IPNLLPMPLPPPHVAVPVLPIPPLPPPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1027
```

```
                    |  |: |  |     :  ||:::|  |
Db    360 ------------VPPPLPPEEP-----QSEDPEEDARLKQLQAAAAHWQQHQQHRVGF 400

Qy    949   LGINGVHLPI-----------STSPNQGFSSPPAVVPPPFAPPPLEPSLSIQEKQGPL 995
            ||   |  |              |:   ::||  :| |     : | |:
Db    401 QYQGIMQKHTQLQQILQQYQQIIQPPPHIQATTPPPGIPPPGVPQGIPPQLT-----  452

Qy    996 RDLFGPCSPNPPTA-------PAPPPPPALPPPL---SGEPKT--PSVESHALTEPMEDK 1043
                :| || :       |  | |  || |:   |  | |  | ::|   :| :  |
Db    453 ------AAPVPPASGSQQQQVPEHPRPALLPTPVSPGSAPPTTYHPPLQSAGPSEQVNSK 506

Qy   1044 NISTKLLVPSGS 1055
              :|  :|  |
Db    507 APLSKSALPYSS 518

Search completed: January 10, 2003, 02.10.28
Job time : 65.6477 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 01:07:54 ; Search time 86.6261 Seconds
(without alignments)
3177.594 Million cell updates/sec

Title: US-09-803-126-1
Perfect score: 6991
Sequence: 1 MGSLEQEAEFQAQTEQNFET . . . WQHFSFLGPAQEVPKTEEVP 1337

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

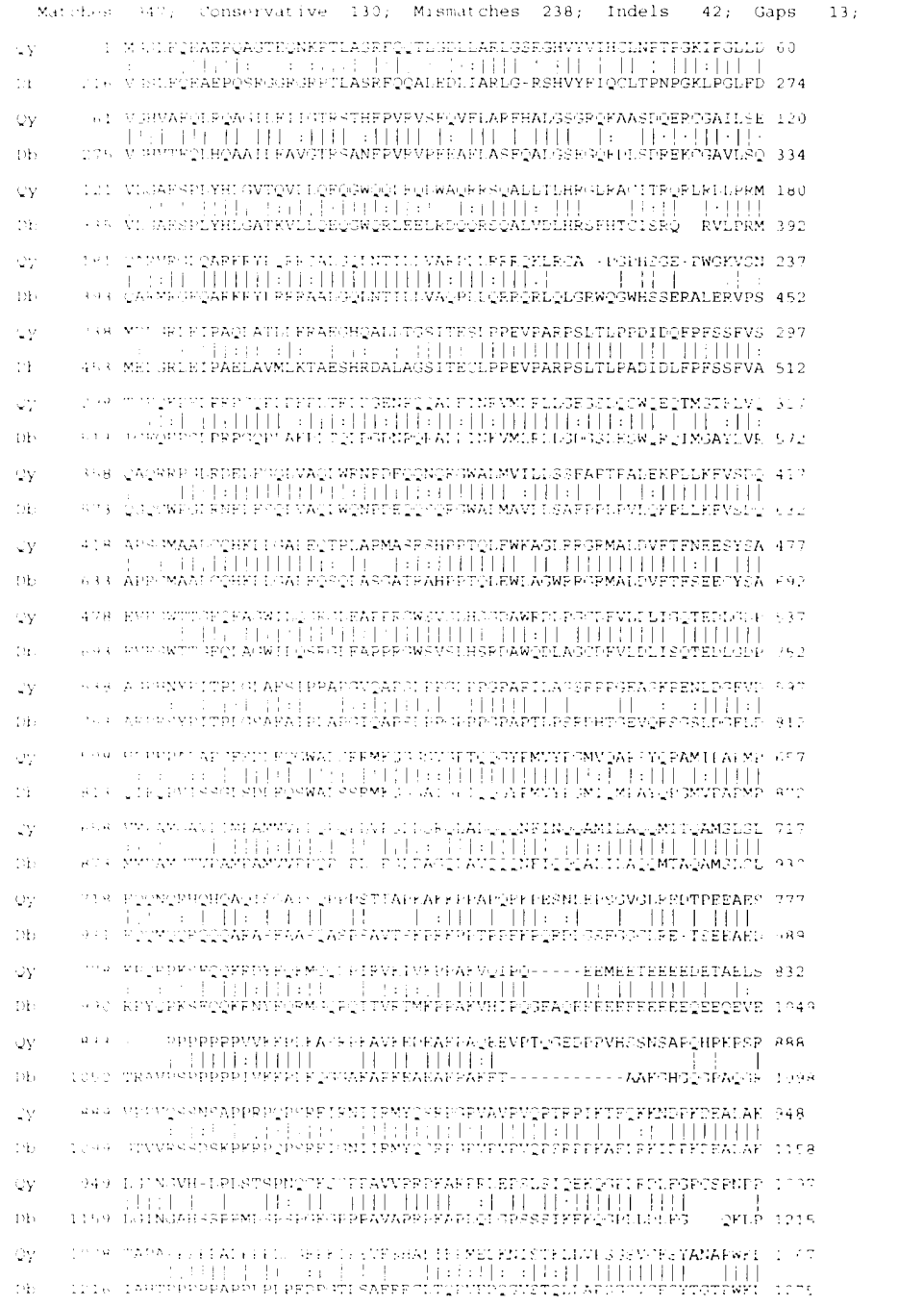
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4775	68.3	1560	4	Q96JP2	Q96jp2 homo sapien
2	900	12.9	2424	5	Q9VZ48	Q9vz48 drosophila
3	512	7.3	2810	5	Q20456	Q20456 caenorhabdi
4	391.5	5.6	2113	11	Q99MZ6	Q99mz6 mus musculu
5	341.5	4.9	2178	13	Q9DGG9	Q9dgg9 brachydanio
6	341.5	4.9	2179	13	Q9DGG8	Q9dgg8 brachydanio
7	339	4.8	1198	10	Q41805	Q41805 zea mays (m
8	338.5	4.8	1315	10	Q9SPM0	Q9spm0 zea mays (m
9	336	4.8	1616	4	O15054	O15054 homo sapien
10	328.5	4.7	5085	11	Q9JKS6	Q9jks6 rattus norv
11	328	4.7	2167	5	Q9V3Z6	Q9v3z6 drosophila
12	328	4.7	3247	12	Q65553	Q65553 bovine herp
13	328	4.7	5147	4	Q9Y6V0	Q9y6v0 homo sapien
14	323	4.6	956	10	Q9LJ64	Q9lj64 arabidopsis
15	317.5	4.5	1236	4	Q9C0I2	Q9c0i2 homo sapien
16	316.5	4.5	2098	5	P91443	P91443 caenorhabdi
17	315.5	4.5	763	2	Q9XDH2	Q9xdh2 mycobacteri
18	311.5	4.5	1045	4	Q9Y6X6	Q9y6x6 homo sapien
19	310.5	4.4	420	5	Q9S547	Q9s547 drosophila
20	310.5	4.4	446	5	Q9VZC2	Q9vzc2 drosophila
21	310	4.4	1611	3	O42854	O42854 schizosacch
22	310	4.4	5038	11	Q9QYX7	Q9qyx7 mus musculu
23	305.5	4.4	1912	11	Q9EPC1	Q9epc1 rattus norv
24	304.5	4.4	409	10	Q9SBM1	Q9sbm1 solanum tub
25	302	4.3	598	16	Q9VXN7	Q9vxn7 mycobacteri
26	300.5	4.3	1175	11	O35126	O35126 mus musculu
27	299	4.3	2121	5	Q9NH54	Q9nh54 drosophila
28	294.5	4.2	2344	5	Q9N3Y8	Q9n3y8 caenorhabdi
29	293	4.2	1052	4	Q96FT1	Q96ft1 homo sapien
30	291	4.2	1435	11	Q9WHF2	Q9whf2 rattus norv
31	290	4.1	1190	4	Q99621	Q99621 homo sapien
32	289	4.1	1175	11	P70206	P70206 mus musculu
33	287	4.1	1123	11	Q9DBD5	Q9dbd5 mus musculu
34	286.5	4.1	1182	4	Q99495	Q99495 homo sapien
35	286.5	4.1	3084	12	Q8UZ11	Q8uz11 pseudorabie
36	286	4.1	1006	11	Q62901	Q62901 rattus norv
37	284	4.1	1154	4	Q9P1P3	Q9p1p3 homo sapien
38	284	4.1	1431	4	Q8WXD9	Q8wxd9 homo sapien
39	282	4.0	1289	10	Q9FLQ7	Q9flq7 arabidopsis
40	281	4.0	5120	13	Q9PU36	Q9pu36 gallus gall
41	280.5	4.0	3164	12	Q69088	Q69088 human herpe
42	278.5	4.0	2067	11	Q9JLI9	Q9jli9 mus musculu
43	276.5	4.0	2280	5	Q9V8E6	Q9v8e6 drosophila
44	276.5	4.0	2302	5	Q9N693	Q9n693 drosophila
45	276.5	4.0	4957	4	O14687	O14687 homo sapien

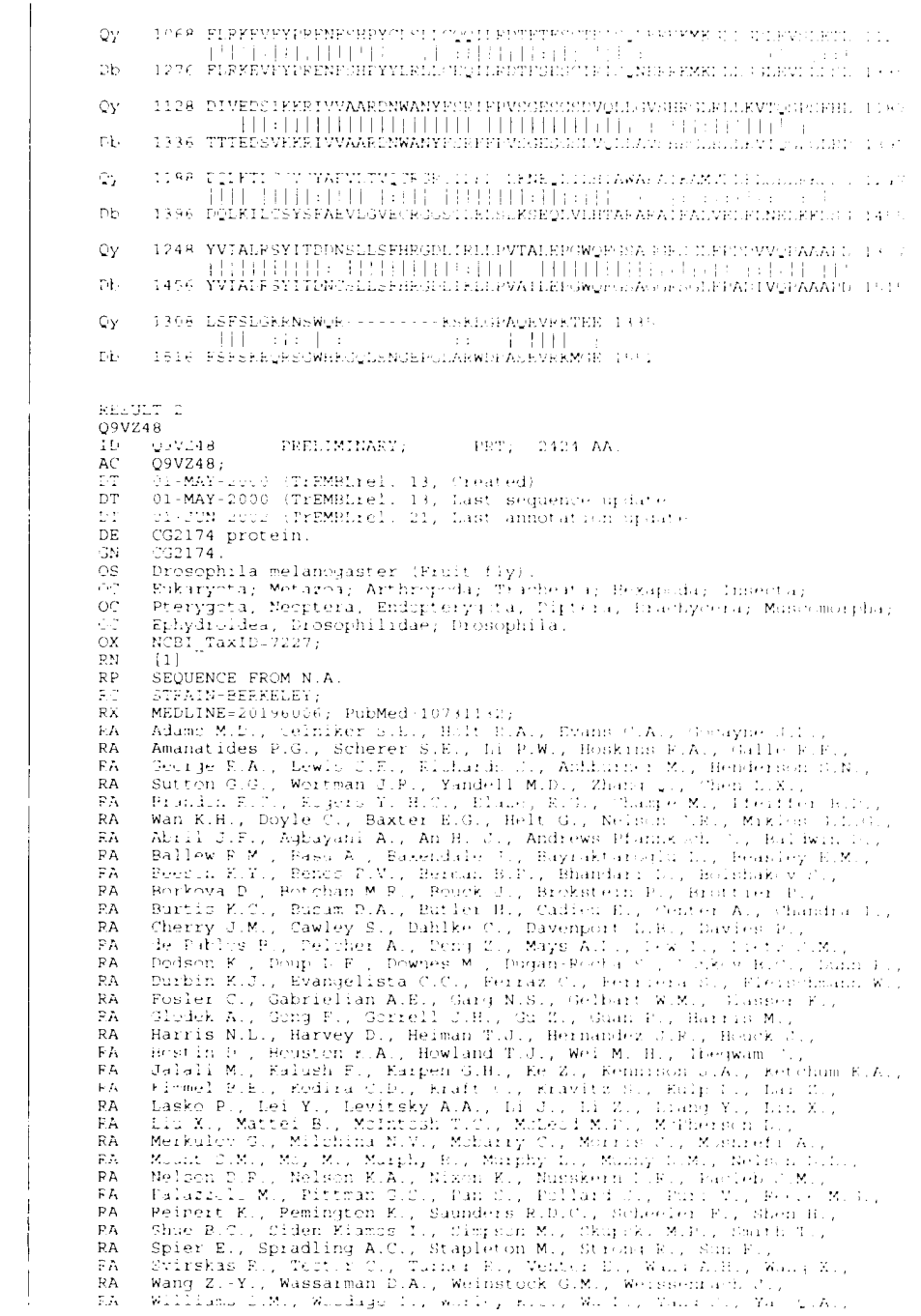
ALIGNMENTS

RESULT 1
Q96JP2
ID Q96JP2 PRELIMINARY; PRT; 1560 AA.
AC Q96JP2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1783 protein (Fragment).
GN KIAA1783.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058686; BAB47412.1; -.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MyTH4; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50002; SH3; 1.
FT NON_TER 1 1
SQ SEQUENCE 1560 AA; 170146 MW; 6A61734C61F4C35A CRC64;

Query Match 68.3%; Score 4775; DB 4; Length 1560;
Best Local Similarity 69.8%; Pred. No. 5.6e-299;

Matches 947; Conservative 130; Mismatches 238; Indels 42; Gaps 13;





RESULT 2
Q9VZ48
ID Q9VZ48 PRELIMINARY; PRT; 2424 AA.
AC Q9VZ48;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN 2002 (TrEMBLrel. 21, Last annotation update)
DE CG2174 protein.
GN CG2174.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota, Neoptera, Endopterygota, Diptera, Brachycera; Muscomorpha;
OC Ephydroidea, Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun F.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

```
RA   Ye J., Yeh R.-F., Zavori J.S., Zhan M., Zhang G., Zhao Q., Cheng L.,
RA   Zhong X.H., Zhong F.N., Zhong W., Zhai X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs P.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003484; AAF47980.1; -.
DR   HSSP; P08799; 1MND.
DR   FlyBase; FBgn0030252; CG2174.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 1.
SQ   SEQUENCE   3454 AA;  387218 MW;  8ADF2AA13F4AA5D8 CRC64;

  Query Match             13.0%;  Score 920;  DB 5;  Length 2424,
  Best Local Similarity   22.6%;  Pred. No. 2.8e-52,
  Matches  415;  Conservative  200;  Mismatches  560;  Indels  586;  Gaps   61;

Qy      19 PTLAGRFQQTLVQLLATLLNSTPHYVYVIKPNNTPTPGKIPGLLLVGSHVAEQLRQAGILEIIG 78
           ||:|:|| :|  ||  :| | | . | |  |    |  | :||| |:|. |
Db     706 PTVAAFFADSLQQLIQSMS-PCHFWFVRCIKPNQEPHALPMDMPCVLQQLRYLGMLDTIQ 764

Qy      79 TRSTHFPVRVSFQVFLARF-HALGSGPQFAAASDQEPCGAILSEV--LGAESPLYHLGVTQ 135
           |  :|||: || |: |: | | |  :  :| | | |  . | | | || |:
Db     765 IPQPGYPVPIEFQHFVERYYHLLPSPLARGTPYPEICRALLEAMPRTGVEGPDYQLGATR 824

Qy     136 VLLQPQNWQQLRLWAQPRSQALLTLHPGLPATITPQPL----PLLPPMQAPVPGLQAPK 191
           | |:|  :|| |  |  |  |  |  |  |  |  |  |    | |||| || |  |:
Db     825 VPIPPALHRALESGPTEPIFEAAVSVQPHVPGMLVPPQLAPPQAAATPLQARWRGQRAQQ 884

Qy     192 RYLQ-RRSAL-------PQLNTIIDVAPPLLPPPQPLPCAPSPHSGEPWGK---VSNMD 239
           || : |: ||       |  |  |  |||| | ||  |  |  |  |  |
Db     885 RYPPIPPGAITAQPIWPGPQAPPPVQQLPSDHPPPQEAPPAAQPAPEAPEAKQAVLEPSQ 944

Qy     240 IGRIPIPAQLATLLEPAEG--------HQALLTGSITESLPPEVPARPS--LTLPPDIDQ 289
           |  |:|||:|| :  :|       |  :|:       || ||  :  :|| |: |
Db     945 LSYLDIPAFIAFIYSKIQGWSPPHGDPHLVKVIGT--------VPGPPSSAVQLPEDLGQ 996

Qy     290 FPPGSFVSTSFQKPFLPPPGQPLDEP-LTPIIGENP--QQALEINRVMLRLLGEGSLQSW 346
           | | | |  |       :|: | |||       |||  . || |  |  |  |  |:|
Db     997 FSPTPFQPVPQNTLPIGPPSPITPFLLTPAASPHQQDFIEALAVFFLILFWSNDFALEGA 1156

Qy     347 QEQTMGTPIVQQAQPPPQLRIPLPSQLVAQL-WPNPPEQQNQPGWALMVILLSSFAPTPA 405
           :|:  :  :: |:|    |||||  ||  |  |  |  |  | | |:  |  |:| |
Db    1157 PPPIQAPIIVHPALSPPQLPIPIIVQLTNQVHQLPPNSGEATPLWQLLSQLCLCIFQPEAA 1216

Qy     406 LEKFLLPPVTPQAPPPMAALTQHPLI----GALPQTPLAPMASPSHPPTQLEWFAGLPPG 461
           | |::|| |:|||  :  |  :||      |  |  | | | | | |||:| |  | :
Db    1217 FSKYLMPPVTPEAPPSIPPLLLPQLQPQQQGGTPSSGAVSAGAGPSFVPAWLEWPAWTPGC 1276

Qy     462 PMALDVPTPNPPSYSAPVPSWTTPQPAGWILQSPGLEAPPPGWSVSLHSGDAWRDLPGC 521
           |||  :  :| | :  |:|||: |  |  |  |  |  ||::  |  |  |  |
Db    1277 DMALPLTLPDEASQTVAVDSWTSCFEAAAAIAVSSLGVAS--PGWTLVLDDGQQLTDSCGL 1334

Qy     522 DFVLDLIGQTE----------------------DLGD----PAGPHNYPITPLGLAESIP 555
           |:|:||| : |                      | |    || |    | |  |  | :|
Db    1335 DYVMDLIAPKELCPAPPAPPSPLLPSGAPPAPTTLPDAVERPAVPPPAPPTSSG-KELVP 1393

Qy     556 ---------      ------    --PAPG--  ---------  --  -------VQAPS   564
                          |:||                                    | |:
Db    1394 PPPPASPPLLSPSSAINPFYPPPPPSPGPSQGSSTSQQAVSPSKSLDDLLAGDIVPVPTDC 1453

Qy     565    --  -  ---------------LPPGLPPGP------------------------ 573
                                     | |   ||
Db    1454 PQQEPIHTIGLSPPPINPFYHSAEPLAFVGPLTGFFYQFSQHAGPPSHAASHGSHSSKYA 1513

Qy     574 ------ ---------- APILASS ---PPTGEASKPENLLGFVLHLFPPALATGPSDLE 612
                             || |||      | | :||   ..| |:| || | |     :||
Db    1414 PKAEYATPSSAMSDTSPAPSLASHVPPVPVPSQAS---PVPQPLPQLPSPVLPGSLDELS 1470

Qy     613 QPWAISPFMKGGG--  ------------SVGFTQQGYFMVY----------------P 649
           .:|.  .:||                  :    |  |:|
Db    1471 PAPSLAASIPGGSYPQPNEAESEIPDLDGYINPIPQPIPLVQNLPKLTGKPQLAAIIKGG 1530

Qy     640 GMVQAPSYQPA-----------MIPAPMPVMPAM---------------------GAV 665
           |:  |  :                :  |:||   :                     |:
Db    1531 GVASKPERPDESEIEDLDYINELPPIPVAEGLPKLTSKEHLAVSIPGGSSTSGNGAT 1590

Qy     666 TTMPAMMVPPQPQPQLVPSLGSPQL----  ALQQQNFINQQAMILAQQMTLQAMQLSIE 718
           | :  :| |  |    |:|  :|:       | |  |  | ||  :|||    ||:  :
Db    1591 FLLHGLMQLPGFTPSGPALYGQLVGPAFLQSAMAQMLQIIQQ  DLAL   NLALQTLLS 1645

Qy     719 QQNQRHQHQAQTGGATSQFP-PSTTAPKAKFPPAPQE------KPESNLEPSGVGLPRDT 771
           ||      |  : :|| || |  :  :  || ||            | :||  |
Db    1646 QQ--  -AAAAAANNSTSPPPAPPILSSLSISPPPPQSPMRMPATRSSLVDPSETRHPLDP 1701

Qy     772 PEEAES---------KPQPPFSPQQKPDYPQ PMGQ---------------------- 797
             |::          ||    ||  . |:|  || |
Db    1702 YGRAPTVPIGPWPWPPQQPQPQTPRPPPAPKMPQHQPKTTPQAIHPQMTSNQGMMPSG 1761

Qy     798 -------------DPIRVKTVKPP---AKVQIPQEEMEET----------EEEEDETAEL 837
                        ...  ||              :| |           : :  || ::
Db    1762 PGAPGASATAIEWRPPPIESPTPPPMPGQLMRSGIRIPTTTTTTISAVNNRIPIPETQDV 1821

Qy     838 SPP--------------------PPPPPVVKKPLKASPPK  .  ... ...   AVKED 886
           |                      |||  | | |  .  :              : | :
Db    1822 SQPMQFVVTTFLAPKSPEIGADPPPPQSVQPLPISSPMPQPLPQVTASHSVPSTVSTKSE 1881

Qy     887 EAEPAQEE-----------------------------------VPTQGE        DPPVH 925
           .  ||: |                                   | ||| |         ||
Db    1882 QPAIAPLEETPKLMLQQQQGDLPASVQSSNQGPGGVVLSHATVPIQIERMLGKLCPPPAP 1941

Qy     926 SSNS---      APQHIPPSPVPPVQPPNSAPPPPQLSPPPINIIPMTQPPPQPVAVIP   987
           |            || | |  ||:  : ||| | |:  :     |  :  |  |
Db    1942 PSGGWPGVLLPPAPSVPAPP--PLIKPPSMAPPAPPPAPQSPPTARSPEPEPDYRTSSSQ 1999

Qy     988 PTP---PIKTPQKPNDPKDPAIAPLGINGVSLPLSTSPNQGPSSPPAVV            1033
           .   |   ...  |       .  |.  || |    :  || |||
Db    2000 VVKEHVPAPIQPQERDTPGAVPQQQMISSHHLHLE-  DHSASSSPAVAAWEQAEKEKSR 2056

Qy     974 PPPKARPPLEPSLSI---QEKQGPL----------------------------RDLPGPC 1002
           | |:| | : | |     | :||                            |:
Db    2057 SPSPSPDPEDYSESVWDPAEVEGPASGSGSEPPPEKPEPEPEPEPLYEIPQVEPEPESHKVY 2116

Qy     1003 SPNPPTA---------------------------------------------------- 1009
            | ||
Db    2117 QPAPPPVIQASMDTTGGHPPPDPTGGLATPPTHMAQKYEHEPPPPSSAPSMPEELPGMP 2176

Qy     1010 -PAPPP---PPALPPPLSGEPKTPSVESHALTEPMEDKNISTKLLVPSGS    VGPSYA 1061
            |||    |  :||| | |                    : : :|: |||       | :|
Db    2177 MTTPPPVIVPAPVPPPSSTSP-------  ---GLGVVSGVLMGSGGGSTACLTYN 2221

Qy     1062 NAPWPLPLPPEVPYPPENPSHPYCLSLLGQTLPDTP-TESTPPSQDPPHPMKGLLGDL 1120
            ||||  |||| | |     |  | ||  |:| | |  |  | ||:  |:      :|
Db    2222 PVPWPLPVPPEVPQPHEPTGPPVALGLPPAQVLGIVPGNTPGLAITPQEPGGALMKLHGH 2281

Qy     1121 PVSLETLD ---------     TVEDSIKKRIVVAARDNWNNYPGRIPPVSGEG  QGPVQ 1165
            ||:||  |  .|. :|  ||| |  ||.|.| |  |  |
Db    2282 GVSVDTLSNFNSGGAGGGGNGGQVPALVPPHLVDMARD WHLYPAPIPAVQGAPLVPDVS 2340

Qy     1166 LLGVSHPGLRLIPVTQSPSFHLDQLKTLCSYSYAEVLTVQCPGR-STLELSLKN-EQLIL 1223
            :|||| || | .         | |  . . |. |. .     |  . |:|:|:| : | |
Db    2341 IMGVSHSGLYIAP------RDADYLIVVQAISPGEIQSAVTLPPPAALQLNLRNGKHLAL 2394

Qy     1224 HTAWAPAIKAMVDLPLSELPK 1244
            | | | ||::||  |: | ||
Db    2395 HAARAAAIQSMVTTPVQEFRK 2415
```

```
RESULT 4
[illegible]
ID   [illegible]      PRELIMINARY;      PRT;  2810 AA.
AC   [illegible];
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F46C4.3 protein.
GN   F46C4.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jotham A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98513613; PubMed=9851916;
RA   [illegible];
RT   "The complete sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z66568; CAA91469.1; -.
DR   HSSP; P13087; 1BR2.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 3.
DR   Pfam; PF03784; MyTH4; 2.
DR   ProDom; PD000855; myosin_head; 1
DR   SMART; SM00295; B41; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   PROSITE; PS50096; IQ; 1.
SQ   SEQUENCE   2810 AA;  323576 MW;  275FF5064B758FB2 CRC64;

  Query Match          7.3%;  Score 512;  DB 5;  Length 2810;
  Best Local Similarity   13.6%;  Pred. No. 6.6e-24;
  Matches  [illegible];  Conservative  243,  Mismatches  982,  Indels  408,  Gaps   61;

Qy      41 PDYQ[illegible]  1 100
Db     825 HVHPVP[illegible]  884

Qy     101 [illegible]  156
Db     885 LFKDIAQCQNEKE---IIQDILQQGQVEYEDLFKIGTEYVFLPEPLAERYDGLQNKICGD 941

Qy     157 ALITHRGLRACITR-  QRLR-LLPRMQAPVRGLQARKRYLQRP    SALGQLNTILL 208
Db     942 AAIIIQKNMKSFVAQKVYKRMRAAIIKLQSCGLRCWKARRDYIIKPEEEMPKAIGR------ 995

Qy     209 VAPTIIRPQKIRCAPGPHCGEPWGFVSNMDLGPIEIPAQLATLIFPAFCHQALLTGSIT 268
Db     996      TMKRNKELDAVHQALLTLNDSQLQQILVSYIDNLLAKPFLEPIE DSDAIETLI 1050

Qy     269 ESLIPEVPAPPSLTLPPDIQQPPSSFVSTsr  RPPLPPPGLPLDSPPLPTLLGE-NPLQA 327
Db    1051 HY   TVPMPNPLQ   NMNPSITLPQFAFENPPTDHIIEPPPPPIMTPPIHPPSYDPPLS 1105

Qy     328 [illegible]  387
Db    1106 VEITPRILLKYMNCIPLTKRLRECLSRYIVQQSIDNIGLRIEILVQTINLNPNQDKTASD 1165

Qy     388 PDWAIEVILSssAs[illegible]  444
Db    1166 DGWGLTPMAIVTPITTENIIPMLIGFPNREVTMPELLFATLQPRL  RITDSEIA    1219

Qy     445 PIHPTQLLDWFASLRSGPMALLVPTPNEECYIAEVELWTTGEQFAGWILQSRGLEAPPRG 504
```

```
Db    1220 RELPTDNLELIATINIFNTVAEINCYGGLMTDVHLGLWNTTGEIARPIIPLKGV GIALG 1279

Qy     505 WSVSLHSGDAWRDLPGTDPVLDLIGQTHDMG PPAPPH    NYIIIIGIGAECIIAFG 564

Pl    1280 WTVPGTTPGPVPAPTPRPIHPVPPQIPGADMTPPPLLMIPPPGIIIPPPPPPPPPVPPEV 1339

Qy     565 VQAPSLPPGLPPGPAPILASSRP                  PIEAGPPPNNLDGPVPHLPE 624

Db    1339 VQEQEIVGQNTNGTIPKAARTYPTPEEWINQPTQRTEGNNTPPCHLPP       DERGI 1393

Qy     625 PAIAPGPSLLEQGWA  LSRRMKGQGSVGPTQL                    GYIMVY 674

Db    1394 ESLVREYNE-EAGYAYTLPRKNK-----PAPLAVELNETLESRELIRIAANPCSPETVR 1447

Qy     675 P  -   - ------ GMVQAP-  --          SYGIAMIIAIMIV      MIAMIA 724

Db    1448 PEPNMPMMETIICHICHILPLIVIIPPTTCPCPLHPITTI MMAAIIPTTTTPLLMIMMMGY 1507

Qy     725 VPTM       PAMMVPPQPQPLNPGLGGPLQLALQGGNPINLQAMILAGLMTTLAMGIAI 777

Db    1508 VPVMMTSSMMTPSMMTP  --SMMP                 LLIAMVPLLMMMLPHPTYY 1561

Qy     778 PQQNQPHQHQAQTSGATSQPPPSTTAPKAKPPAPLPPPPCNCPPCYVLPELTPPAAET 837

Db    1562 PQYPQIPQY----------------QPPPPIICPQS        VPPPVPPMMAI 1607

Qy     838 PPQPPSPPQQKPDY PLPMQLPPIPVPTVPPPAKVQIPQ              PPMPETPP 897

Db    1608 VMHDGHGSTPSPDYPIMPPSPVPSQYVTIP   NMIVIPHPPCVLQPLIAVPLVLPLP 1661

Qy     898 P----------- - -  - -PDETAPLSPPPPIPPVTPP    IIPAPPPA   960

Db    1662 QPAAHPNPNQLANTIPQPIPIPPQYYPIPVLTVPQPVPPVYQTSPTIQPQP PPAPQII 1721

Qy     961  --   ----VKEDEAPPAQEPVPTQGED             PIVENQN      QAPQH 1020

Db    1722 RIIGPPHEHPVRPPHSRNGTSNESTSGDLQMNTPTRGPPPIPPPLNN NYNYPPQIVY 1781

Qy    1021 PPPSP     --   VPPVQPSNQAPPPQ          PPPPIP          NIIPMY 1079

Db    1782 MMIVQMNGHGEMILLSVASPPPAQSPHHTHLPGPPPPCYL TPPTMAPPPPPVPIPPP 1841

Qy    1080 QSPPGPVAVPVQPT-  -  -  -  -RPIKTPQK  - - KNGPKPEALAKLSDINPVHLPI 1138

Db    1842 MGPETPDAM VPGTYVPTHGAQTVVPQTGNTTPPGALPKGSAAIAMPMPLIAPQIPA 1901

Qy    1139 SILPNQAPLPP       AVVPPPAAPPP             LPPLT                 1191

Db    1902 AVPTQNGNVVPPPPGGPPAPGPAPTNGPPPLNPLPLPLNTTPPAVPNIPPNPPYLPTPPP 1961

Qy    1192 SIQPPQSPLR--------  DLPGPGSPNPPIAPAPPPPPAPPPP P PPIPVPPHA 1251

Db    1962 NLDPQPAVIPPPPLIAPNAINMLSDPLPSPVPQPTPPPP   PPPVPPPWVIPDGVPPGH 2021

Qy    1252 LTPPMPPPNISTPLLVP         SGSVVPPGYANAPWPLPLPPPVPVPPLNPGHPVPLP 1311

Db    2022 VPLQGRIPSSPPPPPLVPLPIPPPIVVAPPPAVPPPVPAIWPLIIPPPPMPIPPPPPPPP 2081

Qy    1312 LLPQLLPIPIPIPSSLIPISQVPPPPPMPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1371

Db    2082 QVPAQIVPEDCPKSYPYPIPPQDPPQVETVLPQYPIPPPSPLNNQPNIPPCVPVAVIPLAP 2141

Qy    1372 NWANYPGRIPPV     SGPIGPTVQLLPVGPPPPLPLLPPVTQPLPPPLPLSTPLPPPYAP 1431

Db    2142 LWPLYPNQVTEVVEKRPDEJVQTI  PAISEHGIKL  IVHTPPLLPNIPKIQPPPPPET 2201

Qy    1432 VLIVQPRGRSTLELSLRNEQ      PPLPLHAWAPAIPAMVIPPIPPPPPPPPPYYIALP 2491

Db    2202 IAPVPLEANDILQVHVPPPEPPPNAYGAVRIKINQAIQIPKTLPPPPPPPPPPPPPPPP 2261

Qy    1492 SYITPDNSLLSPHRGDLIRILPVTALEPGWQPGSAGPGSP 1531
```

```
Db   2217 KFV----------  IAPTGPFTPPVGNWLYGKIENFFG 2243


RESULT 4
Q99MZ6
ID   Q99MZ6      PRELIMINARY;      PRT;  2113 AA.
AC   Q99MZ6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Myosin-VIIb.
GN   MYO7B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 21295050; PubMed=11401444;
RA   Chen Z.-Y., Hasson T., Zhang D.-S., Schwender B.J., Derfler B.H.,
RA   Mooseker M.S., Corey D.P.;
RT   "Myosin VIIb, a novel unconventional myosin, is a constituent of
RT   microvilli in transporting epithelia.";
RL   Genomics 72:285-296(2001).
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AF242411; AAK28339.1; -.
DR   HSSP; P08799; 1MND.
DR   MGD; MGI:107709; Myo7b.
DR   InterPro; IPR002106; AATRNA_ligaseII.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000871; Beta_lactamase_A.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR000508; SigPTase.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00744; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS00146; BETA_LACTAMASE_A; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS00741; SPASE_I_3; UNKNOWN_1.
KW   SH3 domain.
SQ   SEQUENCE   2113 AA,  [illegible] MW,  [illegible] CRC64;

  Query Match          5.5%;  Score 381.5;  DB 11;  Length 2113;
  Best Local Similarity  [illegible]%;  Pred. No. 1.2e-15;
  Matches  518;  Conservative  201;  Mismatches  542;  Indels  585;  Gaps  68;

Qy      29 TLASRPQQTLGDLLARLGSPGHVYYIHCLNPTPGKIPGLLDVGHVAEQLRQAGILEIIGT 79
           ||||:|:|:| | |:|:  |  :|: |  | | | |     .||| :|..| :
Db     628 TLASQRPGSL PQIMKILTNQGFYFVPCIKPNEYKKELLFDRELCIQQLRYSGMMETVHI 686

Qy      80 RSTHEPVRVSPQVFLARPHAL--------QSGPQFAASDQERQGAILSEVLGAESFLY 129
           | : ||:| :|  |  ||| |           ||     : |       || :   :
Db     687 RPSSPPIRYIPDEFSQRFRVLLIGFERMQFQNKPRQMTLHIADLC-------LGTDKD-W 738

Qy     130 HLGVTQVFIQRQRWQQLQLWARRPSQAL ----  --- - ---- -----  --LTLH 162
           :| |:: |::      ||     ||||||                      :||
Db     739 PQPPTPIPLMTHLQTVIE  ---IPSSQALLGAATRIQPVLPGHFTPEEFLFQPRAAVTLQ 794

Qy     163 AGLRACITRQRLRLL----PRMQA----  --    ---- RVRGLQARKP-YLQPPSALG 201
           | |    |:  :|:     |:||              |: |||| | || |:
```

```
Db     795 AGWPGYSQPFNFFLILVGFFPLQAIAPSHLLMRIFQAVEQSIVQLSAFGLSLVLQGTQA 854

Qy     202 QLNTILLV  -APPLLPPP--- QKLPCAPGPH---SGEFWGFVSNMFLSRLEIPALL   257
           :      :  ||  . |:     ||    : ||   :|:  :  |   |    :
Db     855 KRRAVVITQAHARGMVVPPSYWQQK---STGPQVILAKEPKAQVAVHEPKRKSIYDTVTD 911

Qy     258 ATLLERAEGHQALLTGSITESLPP---------------EVPARPSLTLPPD----IDQF 298
             ..|:  |    : |       |                ::   | : :|  :    : ::
Db     912 TAMVEKVFGFLPAMIGGQPGPAPTPFEDLEVKTQKLHEVDLDTVPMMAMPLEEVDDLADV 971

Qy     299 PFSSPVSTSFQKPPLPPPQQ PLDEPLTFLDGENPQQ ALEINFVMLFLLGE-------  340
           |  |  | |||         | ||  ||    : :     ||::  ::|| :|:
Db     972 TFPKPAVTYFQKSASHTHIQKPLSYPLLYHENDTDHSAALDVWIIILRFMGDLPDPVVTG 1031

Qy     341 -------------------------GSLQSWQEQTMGT------------------  353
                                      | |   :  |
Db    1032 PNSLTGSSVMFQIHDKLGFDSVTQHIFSSQVASQLNFGEEAFFFDGFISGFDMSNLEKVH 1091

Qy     354 FLVCQAQPPPGLRDELFSQLVAQLWRNPDEQGNQRGWALMVILLSGFAFTPALEKFILKF 413
           |:|  |  |||||||--  |:  ||  |     |||  |: : |  | |:    | || |
Db    1092 FIVGYAIMPPGLFDRFIYCQICFQLSENTYTSGFAFGWTLLSLLLDPFFSEFFMKYLLNY 1151

Qy     414 VSDQAPSGMAALCQHKLLGALFQTPLAPMASPSHPPTQLFWKACLPPGPMALDVFTFNEF 473
           :| | |      |  :|     |  |   |: ||| || :|    :  :  :
Db    1152 IS QGPPSYGPFCAEFL  ---QFTFA-NGVPAEPPTWLELQAVFSPFHIFIQVILATP 1207

Qy     474 SYSAEVESWTTGEQFAGWILQSFGLEAPPPGWSVSLHSGDA-WEDLPGCDFVLDLIGQTE 532
           | :  |:| :|  :    : | :||      |:|: :   |  |   ||| ::| : | |
Db    1208 SLTISVDSASTSFEICQHVAQKGLF-DNLGFSIQVAVYTFFWSIGSSIHLMDAVAQCE 1263

Qy     533 DLG----------------------  -- DP  ----          AGPHNY  -  - 563
           |                           ||                  :| :|:
Db    1264 QLAPERGESIPQAPWPIYFPKEFFTPWHISQFTPVSTFLIVHIQVLFCVWSGETNPEEHEE 1323

Qy     564 ----------------------------------------------------------P 564
Db    1324 LVELLARHCYVQLGATVKSNAVQELLFSCVPSFLYPTKSFFFWASLVTAAHAFAQYTQSE 1383

Qy     565 ITPLGLAESIPPAPGVQADSL ------   --PPGLPPGFAPILASFFPPGFAPMIEMLL  623
           |||  : |    |  :  | |              | ||     :||              |
Db    1384 ATPLAVREQTVEAARLLWPLLFSPLFEVTTLSGPPLPKTQL-VIAI------      NWK 1431

Qy     624 G -FVDHLFEFALAPGFSDLEQGWALSFFMSGGS--VSPTQLSYFMVVIAMVLAPLYLF 680
           |  |:|      |   |:: |  :|  ||    |  :| | | |
Db    1432 GMYFLDQKERTLLGLSFAEV-MGLVANRDAPGGKKLLLATLQEEYEFVSPSSV-----    1484

Qy     681 AMIPAPMFVMPAMGAVPTMPAMMVPPQPQPLVPSLDSPQ---LALQQQN     -  PI 732
                      |  | |:      |  |   |  :|||  :      |
Db    1485 --------------AIAEMVAL--------FLGGLKERSVFAMALQDPPATPDITLLPFK 1531

Qy     733 NQQAMILAQQMTTQAMSLSLEQQNQPHQHQAQTSGATSQPP-------PSTTAP------  787
           :|| ::   |      | | |      :| | |    |  | |  |
Db    1532 KGDLLILTFKQGLLASENWALGQNDF--   --TGPTGLVPTACLYTIPCVIFLCTQLLS 1589

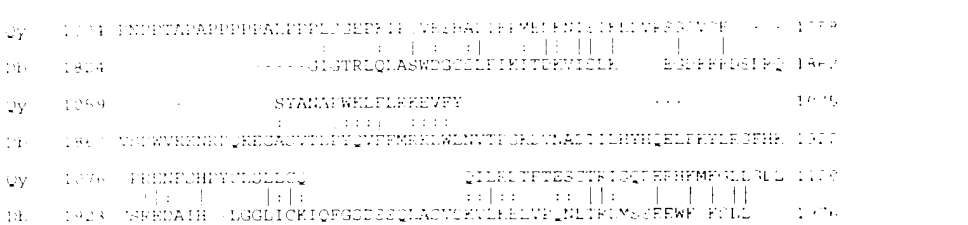
Qy     788 ----  -----   ---KAKKPPAPQEKPRS--NLEPSGVGLPEITPEEAFSFPQPPESPQ  846
                         :| :||   :  ||  ||              :|  |:  |  :
Db    1590 LLAMSPEKRKLAAQEVRALEPPLFDQLTESPYTLPEFSYQFPPAPEKETISRAAMPMA-   1647

Qy     847 QKPDYFQKMGQDPIR --VKTVKPPAKVQ      ------      --IFQEEMEETEEEEDE 897
           : | :      :|:|   :|:|   ||::                    |  :  : | |:
Db    1648 RSRGHLWAYSPEPLRQPLLKSVHDKAKLPDAACQIFLAILEYTGDYPSRQSWHSLELTDQ 1707

Qy     898 --TAELSPPPPPPTVVKFPLKASPPPAVPEDEAPPAQEEVITQDEEPEVRGSWRAPQHPK 955
           :  | |      :  :|      :::  |     |  |  ||          |  |
Db    1708 MFSLALQDPALQLEIYCQILKQLTHNSIPSSEEPAWQLLWLCTGLFPP  GKTILPHAQV 1765

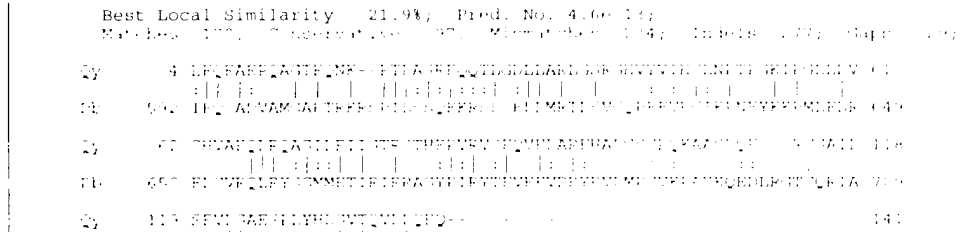
Qy     956 PSFVPPVQSGNSAPPRDQPSPEIPNIIPMYLPEGEVAVEVQFTPEIKTPIPFNIFKQFA 1015
           :  |      |     ||  :  ::|:  :  |  | |:                  ::
Db    1766 -----FIDSPKKKPLALDCSPPLHPVLPVGPPKQPPHDVEVKAA- --          PQM 1820

Qy     1016 LAFLGINGVHLFLSTSENQ--GKSSPPAVVEPPEAPPELEPSLSIPEPQQPLEFLPSPQS 1073
            :.||  :  |.:|  ||  :   | |||                        :|||   |
Db    1821 VSKLH-HEVYLPNDTSKSMEVGSSS----------------------------RVRDL---CE 1863
```



RESULT 5
Q9DGG8
ID Q9DGG8 PRELIMINARY; PRT; 2179 AA.
AC Q9DGG8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Myosin VIIA.
GN MYO7A.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUE;
RX MEDLINE=21411610; PubMed=11958658;
RA Ernest S., Rauch G.J., Haffter P., Geisler R., Petit C., Nicolson T.;
RT "Mariner is defective in myosin VIIA: a zebrafish model for human
RT hereditary deafness.";
RL Hum. Mol. Genet. 9:2189-2196(2000).
DR EMBL; AJ404001; CAC05418.1; -.
DR HSSP; P18799; 1MND.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MyTH4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00612; IQ; 5.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MyTH4; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00295; B41; 2.
DR SMART; SM00015; IQ; 4.
DR SMART; SM00242; MYSc; 1.
DR SMART; SM00139; MyTH4; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50057; BAND_4_1_2; 2.
DR PROSITE; PS50096; IQ; 2.
DR PROSITE; PS50002; SH3; 1.
SQ SEQUENCE 2179 AA; 251632 MW; F93ABB8128DEBBFB CRC64;

Query Match 4.9%; Score 341.5; DB 13; Length 2179;



RESULT 6
Q9DGG8
ID Q9DGG8 PRELIMINARY; PRT; 2179 AA.
AC Q9DGG8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Myosin VIIA.
GN MYO7A.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUEBINGEN;

```
RX   MEDLINE=20414642; PubMed=10958658;
RA   Ernest S., Rauch G.J., Haffter P., Geisler R., Petit C., Nicolson T.;
RT   "Mariner is defective in myosin VIIA: a zebrafish model for human
RT   hereditary deafness.";
RL   Hum. Mol. Genet. 9:2199-2206(2000).
DR   EMBL; AJ404002; CAC59419.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF00612; IQ; 5.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000365; myosin_head; 1.
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 4.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50057; BAND_4.1_3; 2.
DR   PROSITE; PS50096; IQ; 2.
DR   PROSITE; PS50002; SH3; 1.
SQ   SEQUENCE   2179 AA;  251655 MW;  FF440D1EFD588F76 CRC64;

  Query Match             4.9%;  Score 341.5;  DB 13;  Length 2179;
  Best Local Similarity   21.9%;  Pred. No. 4.6e-13;
  Matches  170;  Conservative   97;  Mismatches  234;  Indels  277;  Gaps   29;

Qy       4 LPLEAEPQAGTPQNK--PILASPFQQQTLGDLIAPLGSPGHVYVIHCLNPTPGKIPGLLDV 61
           :|| |:   | |  |  |||:|:|:::| :|| |  |     : :|: |   | | | |
Db     644 IPQ-ADVAMGAETPRPSPTLSSQPKPSL-ELLMPTLSVCQPFFVPCIKPNEYKKPMLFDR 699

Qy      62 GHVAEQLRQAGILEIIGTKGIHPPVRVSPQVFLAKFHALDSGRQKAASDQE---RCGAIL 118
           ||| :|::| |  |   :|:| :|  |: |:  |  | : |   ::   |  |
Db     700 EIGVRQLPYSGMMETIPIPPAGVPIPYTPVEPVDPYEVLMPGVKPAYKQEDLRGTCQRIA 759

Qy     119 SEVLGAESPLYHLGVIQVLLQEQ---------------------- ---- --------- 141
           ||| :   : :| |::  |::
Db     760 PAVIAPRDDD WQMGRTKIFLKDHDMLLEIPPPKAITDFVILIQFVVPGPKDRSNPLKMP 819

Qy     142     ---------------------GWQQLEQLWAQRRSQALLTLHPGLRACITPQPLPLL 177
                                    |: :|: |:  |:      | |       : |||: |
Db     820 KGAMLIQFTWFGYTCPDVDAMESARFPLLALIESRE  LYQTYH---  -VAPQPIMI  879

Qy     178 PRMQAPVPGLQAPFPYLQPPSAL  GQLNTILLVAPPLLPP------- ----- ----R 217
           | | ||   |: :  |  |:    |  |  ::|| | :|
Db     880 --FQGPCPGPLVPPAPPHPLWAVITIQAYTPQMIAPFLYPFLPGEYPPPLEAPKLPLAPE 937

Qy     218 QKLPCAPGPHSPPFWGPVSNMGLGPLEIPAQLA ---  TLLERAEGHQAL---------- 262
           ||||          :   :  |   ||  ||||          || |  : :
Db     938 IPLPNIMAPAPAPE  PAPPMHLPPL --ALAPPIAPPPPPPPPPQEAPPKMEMLDMEPA 992

Qy     263 -------------LTGSI--TESLPPEVPARP----------------  -SLTLP--- 284
                        . | :  | | | :    |                  || ||
Db     993 RQEPVNDGEMVDPMPGPLGTTNSFPGQEGQAPAGPEDLEPTHPPLEVEDLDESLPLPEDD 1052

Qy     285  PDIDQPPPSSPVSTSPQ----KPPLPRPGQPLDEPLTRLDGENPQ-QALEINPVMLPLL 338
            |: :: |: | :| ||        :: |   || :||   : |   |  || :   :|| :
Db    1053 LEDLSEYKPAAPSATYPQGTSTHTYMRR---PLKQPLLPHEDEGDQLAALAVWITVLPPM 1109

Qy     339 G------------EGS--------------------LQSWQEQTMGT------------ 353
           |            :||                    ||: | :   |
Db    1110 GDLPEPKYHTAISDGSEFTPVMTKIYETLGPKTYKPRELQALQGEGENTNIESHKKJSVRH 1169

Qy     354 ---------------------------------------------FLVQQAQPPPGLRD 367
                                                        |::     ||||||
Db    1170 KLVSLTLPKKSKITEEVTPPLNDGEYTVHGNPMLEPPTSNLEFLHPIIGNGILRPGLRD 1229

Qy     368 ELFSQLVAQLWRNPDEQQNQRGWALMVILLSSFAPTPAIEKPLLEFVSDQAPSGMAALCQ 427
           |:: |:  || :|| :  : |||  || : :  |||:     | |  |:| : | | |  |:
Db    1230 EIYCQICKQLNQNPSKSSHAPGWILMSLCVGCFAPSFKFVKYLENFIS EGPFGYAPYCE 1289

Qy     428 HKLLGALEQTPLAPMASPSHPPTQLEWKAGLPPGPMALDVPTPNEESYSAEVESWTTGEQ 487
           :|              :|: ||: || :|   :  : | |   :  : :   :| ||  ::
Db    1290 ERLPPTPVN- ----GTPTQPPSWLELQATKSKPPIMLPVTPMSQTTYTLLTDCATTAKE 1343

Qy     488 PAGWI---  LQSPGLEAPPPGWSVSLHSGDAWPPL-PGCDPVLDLLRQTPDLGDPAG 539
                 :    || |       |:|: :   |    |  | | | |:| : | |      |
Db    1344 LCNALSDKISLQDP-  ---PGPSLYTALPTPVSSLTGGNDHVMDAVSQCEQYAPLQG 1394


RESULT 7
Q41805
ID   Q41805      PRELIMINARY;      PRT;  1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Extensin-like protein precursor.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta; Magnoliophyta, Liliopsida, Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN,
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RT   "Pex genes: pollen-specific genes with extensin-like domains.";
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003591; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
SQ   SEQUENCE   1188 AA;  120981 MW;  2C77C7F8D7130149 CRC64;

  Query Match             4.8%;  Score 339;  DB 10;  Length 1188;
  Best Local Similarity   25.8%;  Pred. No. 2.9e-13;
  Matches  150;  Conservative   80;  Mismatches  258;  Indels  124;  Gaps   26;

Qy     540 PHNYFITPLGLAESIPFAPGVQAPSLPPGLPP--GPAPILASSPPPGEASKPENLPSPVP 597
           |  ||  |  :  | ||    :|: | |  ||    | |     |:||     ||
Db     663 FEEYPTTPTSVKSSPFPEKSLPFPTLIPSPFPQEKPTPFSTTSKPFSSFEKPSPPK---- 722

Qy     598 HLFEPALAPGFSDLEQGWALSPFMFGGGSVGPTQQGYPMVYPGMVQAPGYQ    PAMIP 653
           |  :|      |    |       |  ||    |     .  ||.    ||  :
Db     723  -EPVSSP--  PQTPKSSPPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPAPLS 774

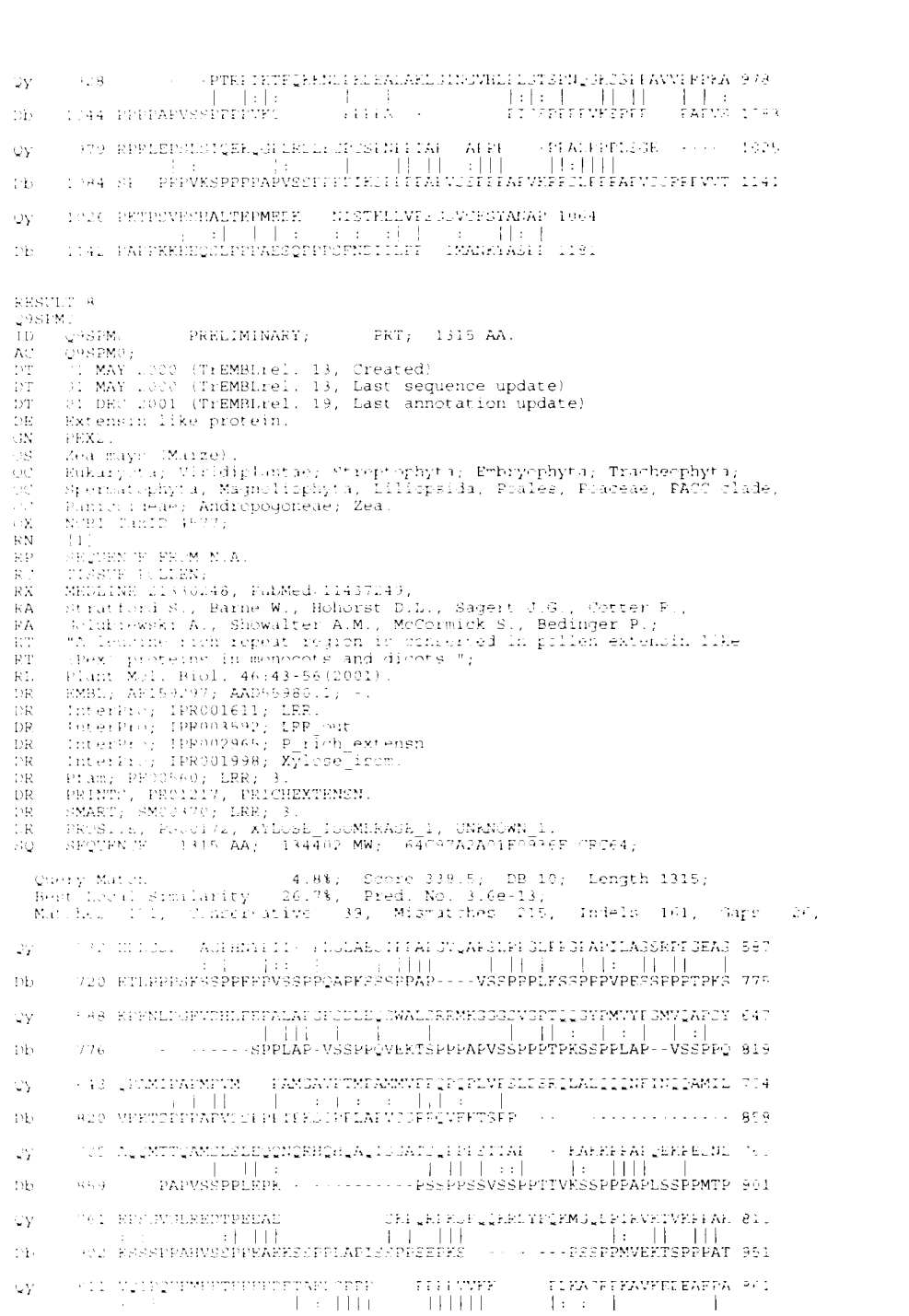
Qy     654 APMPVMPAMGAVPTMPAMMVPPQPQ---PLVPSLDSPQLALQQQNPFINQQAMILAQQMTT 710
           :| |      : |  ;    | | |:   | |  |     |  |  |  |  .  || . :
Db     775 SPFPAPQVKSSPFPVQVSSPPPAPKSSPPLAPVSSPPQVEKTSPPPAPLSSPPLAPKSSP 834

Qy     711 QAMSLSLEQQNQRHQHQAQTSGATSQPPPSTTA-----PKAKPPAPQEKPESNLEPSGV 765
           : :|          :      |||:  :     ||   |||     |   ::||
Db     835 PHVVVS--------  --SPPPVVKSSPPPAPVSSPPLTPKPASPPAHVSSPPEVVKPS-- 881

Qy     766 GLPEDTPEEAESKPQPPKSFQQKRDYFQFMGQDPIRVKTVKPPAKVQIFQEEMEETEEEE 825
           |     |  |  |||          :   |  ||:  |||  |  |
Db     882 - TPPAPTTVISPPSEPKSSPPI--  TFVSLPPFIVKSSPPPAMVSSP------ -- - 924

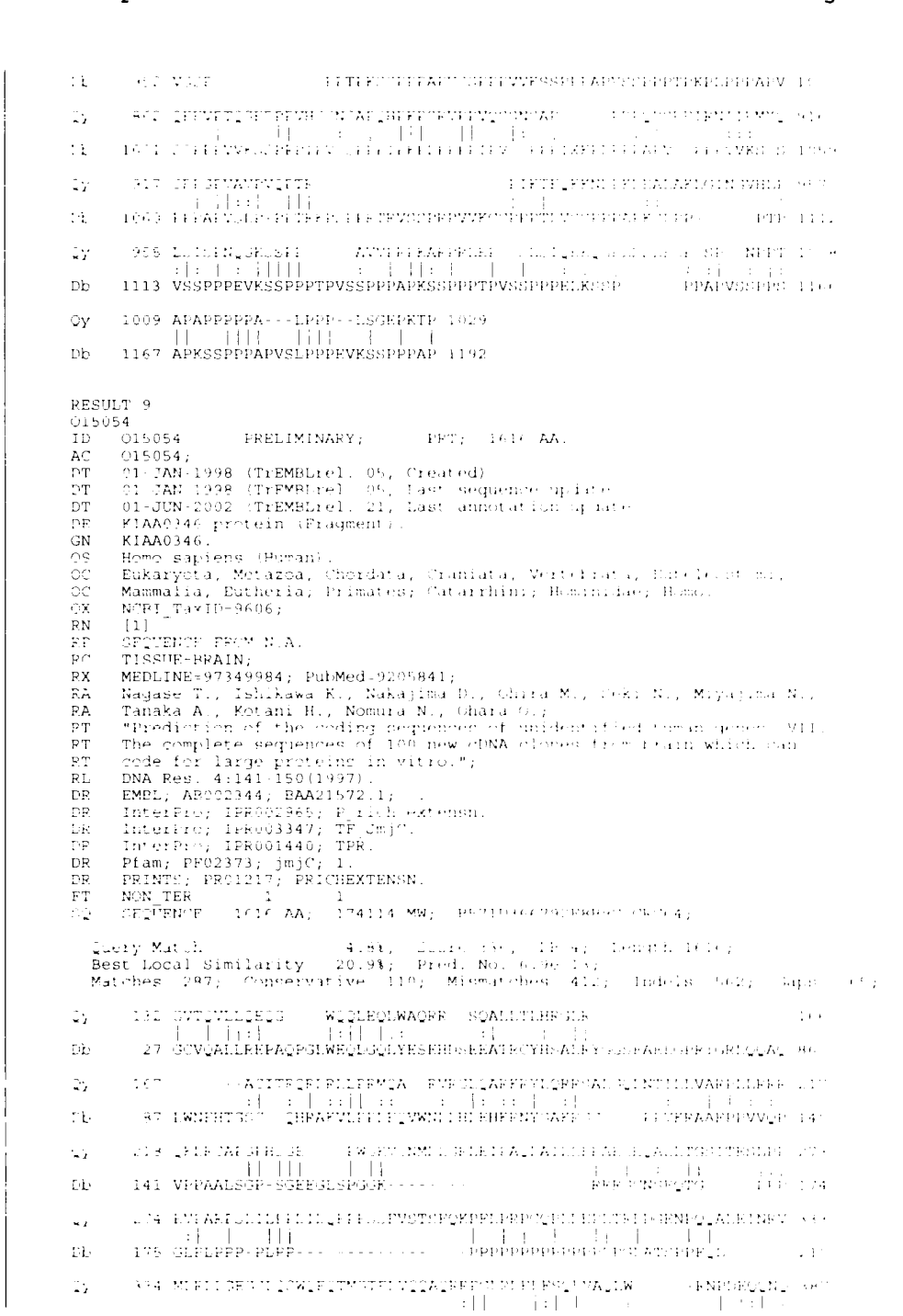
Qy     826 DETAELSPPP----PPPPVVKK---PLKASRPKAVKEDEAEPAQEEVPTQ- GEDPPVHS 876
           | : ||||     |||  ||    |  | | | :    ||   :|      ||
Db     925 PMTPKSSPPPVVVSSPPPTVKSSPPPAPVSSPPATPKSSPPPAPVNLPPPEVKGSPPPTP 984

Qy     877 SNSAPQHPKPSRVPPVQSSNSAPPR---PQPSREIRNIIRMYQSRPGPVAV-----PVQ- 927
           :| |  || |  ||    |:  ||    | |   : :     :| | | |     ||:
Db     985 VSSPPPAPK-SSPPPAPMSSPPPPEVKGPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKS 1043
```



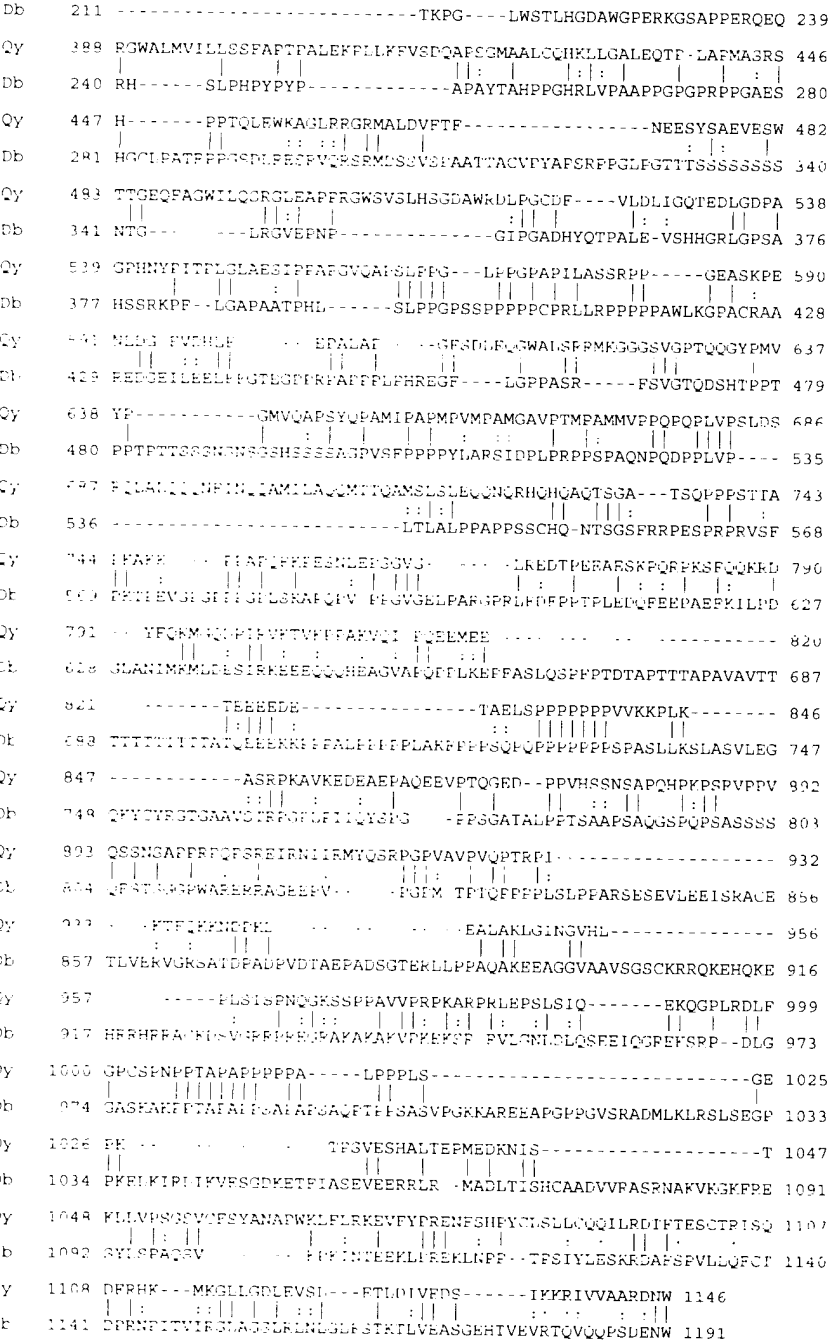
RESULT 8
Q9SPM0
ID Q9SPM0 PRELIMINARY; PRT; 1315 AA.
AC Q9SPM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Extensin like protein.
GN PEXL.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX MEDLINE=21160346; PubMed=11437243;
RA Stratford S., Barne W., Hohorst D.L., Sagert J.G., Cotter F.,
RA Schulz-Swaki A., Showalter A.M., McCormick S., Bedinger P.;
RT "A leucine rich repeat region is conserved in pollen extensin like
RT PRPX proteins in monocots and dicots ";
RL Plant Mol. Biol. 46:43-56(2001).
DR EMBL; AF150797; AAD56980.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_cat
DR InterPro; IPR002965; P_rich_extensn
DR InterPro; IPR001998; Xylose_isom.
DR Pfam; PF00560; LRR; 3.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00370; LRR; 3.
DR PROSITE; PS51257; ATLOSB_ISOMERASE_1; UNKNOWN_1.
SQ SEQUENCE 1315 AA; 134402 MW; 64C97A2A01F0356F CRC64;

Query Match 4.8%; Score 339.5; DB 10; Length 1315;
Best Local Similarity 26.7%; Pred. No. 3.6e-13;
Matches 171; Conservative 38; Mismatches 215; Indels 161; Gaps 26;

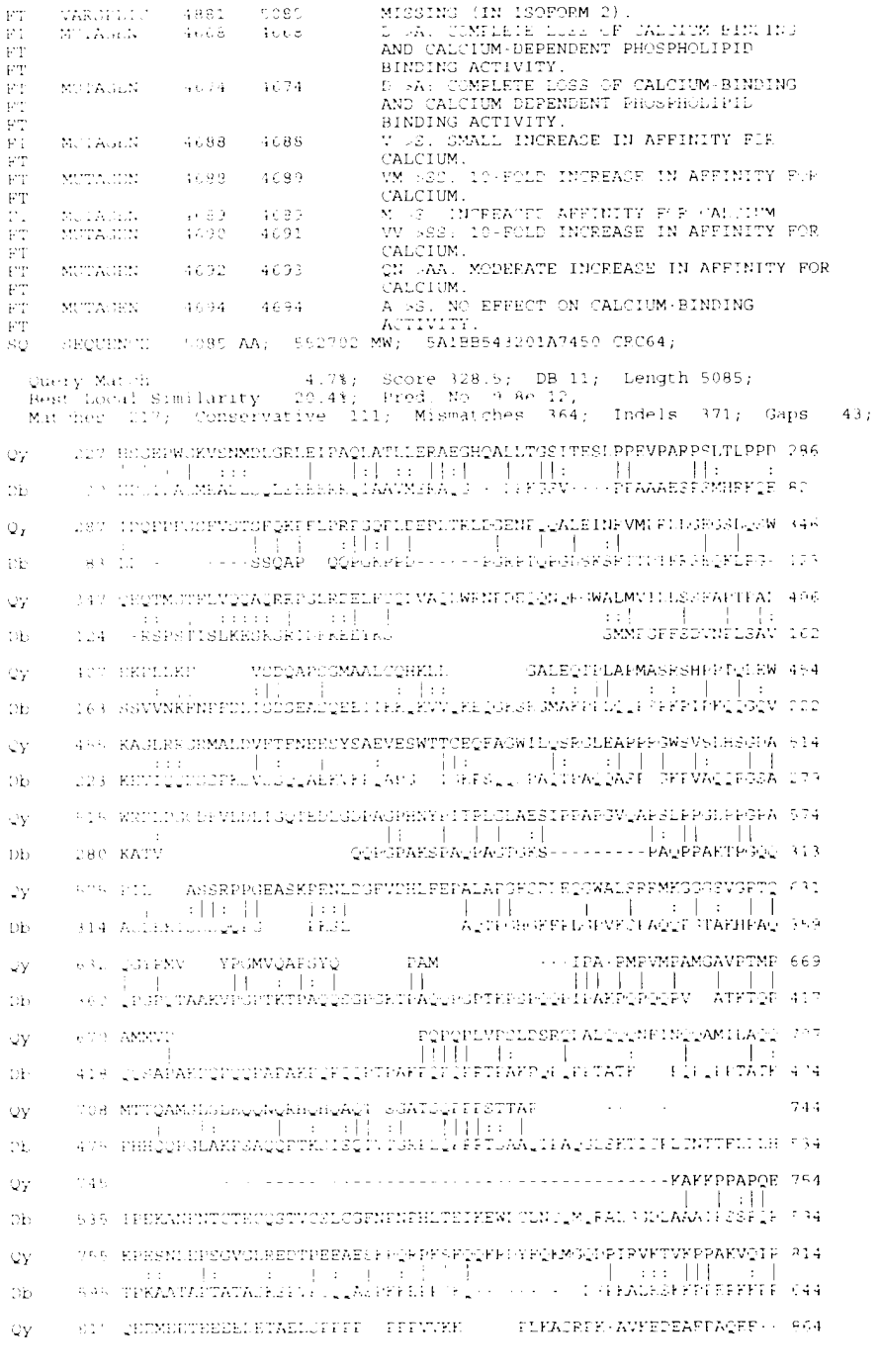


RESULT 9
O15054
ID O15054 PRELIMINARY; PRT; 1610 AA.
AC O15054;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA0346 protein (Fragment).
GN KIAA0346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AP002344; BAA21572.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR Interpro; IPR003347; TF_jmjC.
DR InterPro; IPR001440; TPR.
DR Pfam; PF02373; jmjC; 1.
DR PRINTS; PR01217; PRICHEXTENSN.
FT NON_TER 1 1
SQ SEQUENCE 1610 AA; 174114 MW; [illegible] CRC64;

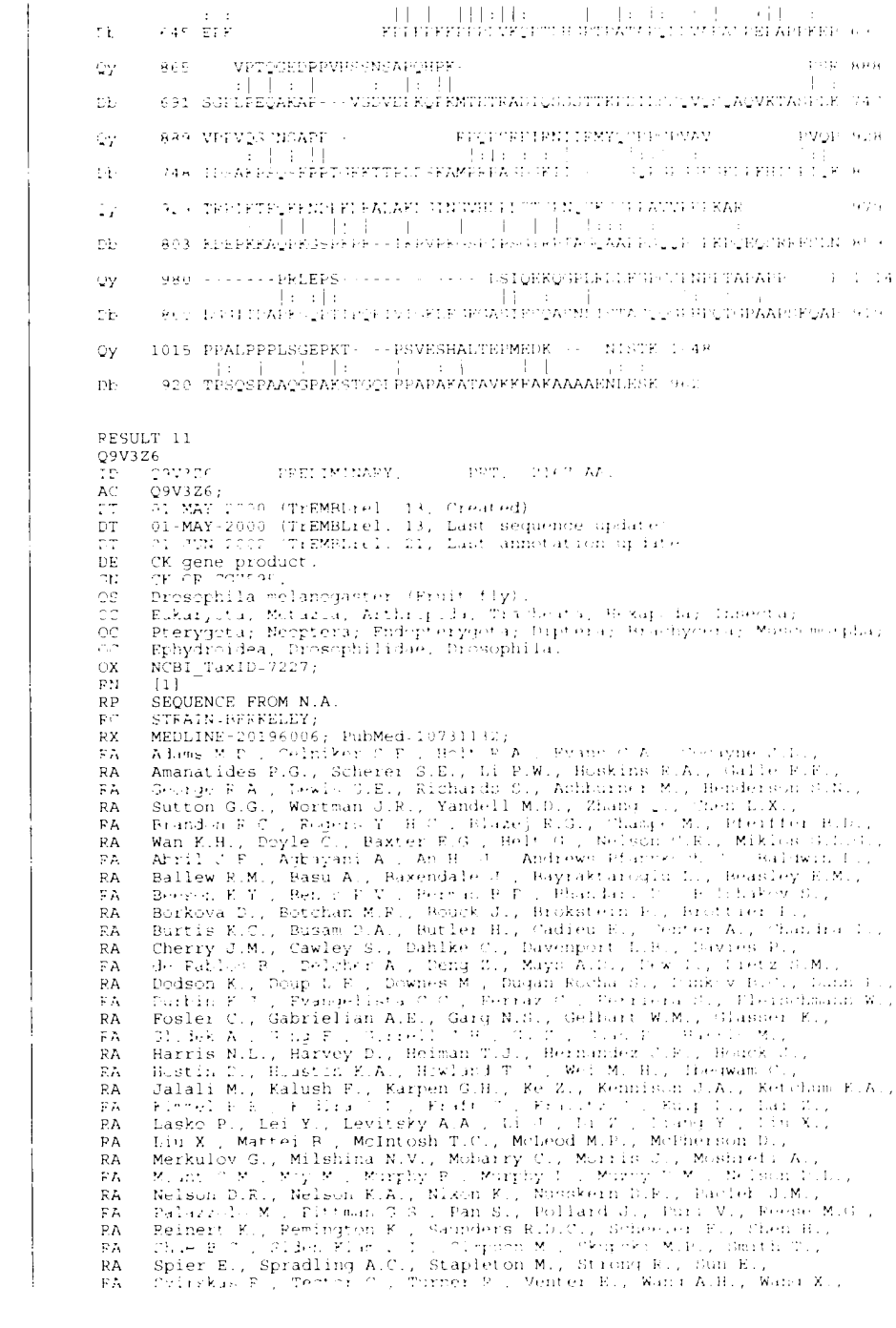
Query Match 4.4%; Score 310.; DB 4; Length 1610;
Best Local Similarity 20.9%; Pred. No. 6.9e-13;
Matches 297; Conservative 110; Mismatches 413; Indels 562; Gaps 44;



```
RESULT 10
Q9JKS6
ID   Q9JKS6      PRELIMINARY;      PRT;  5085 AA.
AC   Q9JKS6; Q9JLT1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN   PCLO.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX   MEDLINE=20170257; PubMed=10707984;
RA   Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA   Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT   "Piccolo, a presynaptic zinc finger protein structurally related to
RT   bassoon.";
RL   Neuron 25:203-214(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP   VAL-4688; MET-4689, VAL-4690, SER-4691, GLU-4692, ASN-4693 AND
RP   ALA-4694.
RX   PubMed=11285225;
RA   Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT   "An unusual C(2)-domain in the active-zone protein piccolo:
RT   implications for Ca(2+) regulation of neurotransmitter release.";
RL   EMBO J. 20:1605-1619(2001).
CC   -!- FUNCTION: MAY ACT AS A SCAFFOLDING PROTEIN INVOLVED IN THE
CC       ORGANIZATION OF SYNAPTIC ACTIVE ZONES AND IN SYNAPTIC VESICLE
CC       TRAFFICKING (BY SIMILARITY).
CC   -!- SUBUNIT: INTERACTS WITH RABAC1/PRA1 AND BASSOON.
CC   -!- SUBCELLULAR LOCATION: CONCENTRATED AT PRESYNAPTIC SITE OF SYNAPTIC
CC       JUNCTIONS.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: C2 DOMAIN 1 IS INVOLVED IN BINDING CALCIUM AND
CC       PHOSPHOLIPIDS. CALCIUM BINDS WITH LOW AFFINITY BUT WITH HIGH
CC       SPECIFICITY AND INDUCES A LARGE CONFORMATIONAL CHANGE.
CC   -!- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AF138799; AAF07822.2; -.
DR   EMBL; AF227534; AAF63196.1; -.
DR   HSSP; P04410; 1A25.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF00168; C2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN      372    491       12 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     523    547       C4-TYPE (POTENTIAL).
FT   ZN_FING    1010   1033       C4 TYPE (POTENTIAL).
FT   DOMAIN     2351   2362       POLY-PRO.
FT   DOMAIN     4442   4536       PDZ.
FT   DOMAIN     4653   4752       C2 DOMAIN 1.
FT   DOMAIN     4968   5059       C2 DOMAIN 2.
FT   VARSPLIC   4876   4880       TYPTN -> SYERK (IN ISOFORM 2).
```



Query Match 4.7%; Score 328.5; DB 11; Length 5085;
Best Local Similarity 20.4%; Pred. No. 9.8e-12;
Matches 217; Conservative 111; Mismatches 364; Indels 371; Gaps 43;



RESULT 11
Q9V3Z6
ID [illegible] PRT; 2147 AA.
AC Q9V3Z6;
DT 31-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 31-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CK gene product.
GN CK OR CG1528.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C.,
RA Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S.,
RA Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M.,
RA Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P.,
RA Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R.,
RA Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H.,
RA Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A.,
RA Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S.,
RA Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z.,
RA Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P.,
RA McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J.,
RA Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M.,
RA Nelson D.L., Nelson K.A., Nixon K., Nusskern D.R.,
RA Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J.,
RA Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F.,
RA Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P.,
RA Smith T., Spier E., Spradling A.C., Stapleton M., Strong R.,
RA Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H.,
RA Wang X.,

```
RA   Wang Z. Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.",
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle P., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9 Mb region of the genome of
RT   Drosophila melanogaster: the Adh region";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Burtenhoff C., Champe M., Chaver C., Chow M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle P., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.P., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Parlieb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang P.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AE003646; AAF53435.1; -.
DR   EMBL; AF0[illegible]41; AAF44915.1; -.
DR   HSSP; P08799; 1MND.
DR   FlyBase; FBgn0000317; ok
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR000048; IQ_region
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR000857; MyTH4.
DR   InterPro; IPR000567; SBP_bac_1.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00612; IQ; 4.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00784; MyTH4; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1
DR   SMART; SM00295; B41; 2.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00139; MyTH4; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50057; BAND_41_3; 2.
DR   PROSITE; PS50096; IQ; 1.
DR   PROSITE; PS01037; SBP_BACTERIAL_1; UNKNOWN_1
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain.
SQ   SEQUENCE   2167 AA;  250307 MW;  3C57E14ADDD89A42 CRC64;

  Query Match             4.7%;  Score 328;  DB 5;  Length 2167;
  Best Local Similarity   21.6%;  Pred. No. 3.4e-12;
  Matches  162;  Conservative   97;  Mismatches  263;  Indels  229;  Gaps   24.

Qy       [illegible] 65
         |: | | | :  |||:::|:::|  |:  | |  :  | |:| |  | | : | |
Db     589 QDIEMGAPTPKPTPTLSTQPPKSLDALMKTLSS-CQPPFIPCIKPNELKPMMPDPGLCC 647

Qy       [illegible] 123
         |||  :|:::| |  |  :|:|  |: |: |  |  |  |  |  |  | : |||
Db     648 RQLRYSGMMETIRIRRAGYPIRHGFPEPVFPYPPLIPGVPPAHRTDCQAATSRICAVVLG 707

Qy     124 AESPLYHLGVTQVILQSQSWLQDLIWAQPFSQALLTLHPGLPACITPQR-LPL------ 176
```

```
Db     708 FSI--YLGHTFVFLKDAHGLFLPQEFLFVLTPFILLQSIRGWVYERRFLSLRAAAIT 765

Qy     177 -----------------LPPMQARVR----------------GLARKE-YLQRRSA 199
                           |:|| :|                  ||||  | || ||
Db     766 VQPFWFGYAQPFPYPNMFVGYMPLQALIPSPVISHFFPHLPGHIVGLQAHAPGYLVPPEY 825

Qy     200 LGQLNTILLVAPPL-----LPPPQKLPCAPGP --------        HSGFFWGK VSN 247
           ::  :: :  :  :     :|| :|||                        | |    : ::
Db     826 GHPMWAVIKIQSHVPPMIAMFPYPKLPLEHFQFAEVIQLPKLEFQFILHPGNFHAPRIAE 885

Qy     248 M---   ---DLGRLEIPAQLATLLERAEGHQALLT-----------GSIIESLPPEVPAR 278
               :| | ||  ||     | | :  ::                 | : |::   :|
Db     886 QHYPIPLHELEPPRIQPLLEN-PPPVPVMNIIINDAAPFLEEFVILGSKLVEAMFDFLPDS 944

Qy     279 PSLTLPP- ---------------------------------DIDQFPFSSPVSTSFQ--   301
               |    |                                 |: :| |  | : | ||
Db     945 SSDAFTPHGGPETSVFNDLFHAQNVNQGDIIAPIHISEDEEPLSEPPFLRPAATTPQQNV 1004

Qy     302 ----------FPFIP----   PPQQPLCEPLTPLCGANPQ        QALEINEVMLRL-  337
                    | ||          | |    : |  |: |:          ::   || ::
Db     1005 NHQVAFFALFHPLLPLHDQGPQLAAQALWITLPFTGDMDEPFYHTMLPMPTISVMSLVT 1064

Qy     338 ------------------------------------------------LGE------   340
                                                            |||
Db     1065 ATLGPNFTPSKEPQEAQIMGLDFDAFIPQFTFSIFHFLVSLTIFPFNFLSEVVPPLQDD 1124

Qy     341 - - GSIQSWIEQTMGT------FLVQVA PFPGLRDELFGQIVAQGWNPEDELHLFGW  390
               | |||  :    .      |::     |  |||| .. |:  ||  ||  :  : |||
Db     1125 EYTADSYQSWLQSPPTSNLEFLHFIIGHGILPAELPDFIYTIIPPGLIWNIIF.[illegible]  .84

Qy     391 ALMVILLSSFAPTPALEKPLLFFVSDQAPSGMAALPQHFLLGALEQTPLAPMASPSHPPT 450
           |: : :  |||:      |  |:  : | | |  |: :|             :|: ||:
Db     1185 ILLSLCVGCFAPSEFFVNYLPAFIP-PIPPGYAIYCFFFLFETFNN    --GTPNQFFS 1237

Qy     451 QLEWKAGLRRGRMALDVFTFNEESYSAEVPSWTTGEQFAGWILQSRGLEAPPRGWSVSLH 510
           || :|  :  : :| :  :  : :  :| || :  :  :  |:   |:|: :
Db     1238 WLELQATKSKKPIMLPITFMDGNTFTLLADSATTARELCNQLSDKIGLK DQFGFGLYIA 1296

Qy     511 SGDAWRDL-PGCDFVLDLIGQTEDLGDPAG 539
           |    |  | | |:| | | |      |
Db     1297 LFDKVSSLGSGGDHVMDAISQCEQYAKEQG 1326


RESULT 12
Q65553
ID   Q65553      PRELIMINARY;      PRT;  3247 AA.
AC   Q65553,
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   UL36.
GN   UL36.
OS   Bovine herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae, Varicellovirus.
OX   NCBI_TaxID=10320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COOPER;
RA   Schwyzer M., Vlcek C., Lowery D.E., Bella I.J., Meyer G., Misra V.,
RA   Thiry E., Paces V.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-179 FROM N.A.
RC   STRAIN=COOPER;
RA   Schwyzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA   LaBoissiere S., Misra V., Vlcek C., Paces V.;
RL   Vet. Microbiol. 0.0-0(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COOPER;
RA   Schwyzer M.,
```

```
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-179 FROM N.A.
RC   STRAIN=KOBA;
RX   MEDLINE=97114286; PubMed=9010999;
RA   Schwyzer M., Ackermann M., Ehlers B., Lowery D.E., Simard C.,
RA   LaBoissiere S., Misra V., Vlcek C., Paces V.;
RT   "Gene contents in a 31-kb segment at the left genome end of bovine
RT   herpesvirus 1.";
RL   Virology 233:67-77(1996).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOBA;
RA   Schwyzer M., Vlcek C., Lowery D.E., Belak L.J., Meyer G., Misra V.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOBA;
RA   Schwyzer M.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z78205; CAB01605.1; -.
DR   EMBL; AJ004801; CAA06097.1; -.
DR   InterPro; IPR005210; Herpes_UL36.
DR   Pfam; PF03586; Herpes_UL36; 1.
SQ   SEQUENCE   3247 AA;  332190 MW;  3AEAA72F8F001F6A CRC64;

  Query Match             4.7%;  Score 328;  DB 12;  Length 3247;
  Best Local Similarity   23.2%;  Pred. No. 5.8e-12;
  Matches  308;  Conservative   90;  Mismatches  348;  Indels  315;  Gaps   53;

Qy     153 WAPPPQALLTLHPCL--PACITPQPIPILPPMQAPVPGLQAPPPYLQPPPALSQLNTIL 209
           :  :    || |   :   ||:| |    | :: |    | | |         |  :| |
Db    2435 WGER-   LIALLENIIENACLTAQ    LEQLGA    LIACPPLPCAPPCLIMLPTH 2382

Qy     210 LVAPPLLRPPQKLPCAPSPHCGPPWPYDNMDCSPIFIPAQIATILEPAPGHQAIITSSI 267
           : : :                | :  :   |  :| |:   | :          ||
Db    2381 QIGDVL                WAPAAAPPLAMLKI-AEDDVLRDLPS-----LTLDY 2418

Qy     268 TLPPPPAPLPTPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 327
           :: :  |  |          |:  :| :   :  | ||    ||
Db    2419 EDALLPAASAADPL------    PTHIISENNVFDFAAASGDFLYAHFFV- -       2459

Qy     328 QALRINDVMLRLLSEGSLG-WGEGTMGTFLVGQAGRPPGLPLPELP-SQLVALGWPNPDEQ 384
                                 :|  |   |   :|    |  :  :  :: ||   | |
Db    2460 --       -  ----  ----PGHATENLFPHAPAAAAGSGGGNRVANARAVAAAAPPDTP- 2499

Qy     385 -NLGGWALMVTLLGGEALTAL    LRLLRPVGL-            ALGGMAALTGR 429
                     |: :| |'|     :| | :  :.:             ||
Db    2500            -ASAPAPAPALAPASAEPALAWGADEWLDDEIGDAYPPAP--------- 2537

Qy     430 KLGPAGPTPLAPMASRSHPPTGLEWPAGLPPGPMALDVPTPNPRSYSAEVESWTTGPQP 488
           :  :  | || |  :: |        :|: ||             : :   |
Db    2538     AHDAYPPAP-AHDAYPPAPAH-----AQGKPAL-------------LAAPAAGAAE 2574

Qy     489 AGWILQSRGL  EAPPRGWSVSLHSGDAWPDLPGGDPVPLQLIQQTPPDLGDPAGPHNYPIT 548
           : ::| |     ||    |: :  || |  |       |:  :  | ||   |
Db    2575 APSVPPAAPPLTPRTGPP   PLMIAALPWERIPPA----SLVSAAPPV--PPGPPIPPAP 2628

Qy     549 PLPLAPPPAPGV  QAPGLPPPLP  PGPAPILASPPPPGPAGPPPPNLPGPVPHLPPP 608
           :   | || :   |  :|| :  | ||| |    ||      |      |  | |
Db    2629 PLP     PPAPPLPPPAPPLPPPAPPLPPPAPPLPPPAPSTAPVPAPPLPPP 2681

Qy     609 ALAPGPSDLPQGWALPPPMPGPLVVPHTPGPVPMVVPP    MVPAPGVPPAMIPAPMPV 668
           :  :   :              :  ||    |:  |     :| ||   || |||| |
Db    2682 ALTPALTP ------------APTPAPT-PAPPLPLPAPITVLVPAPVPAPAPIPAPAPT 2726

Qy     669 MPAMGAAPGMPMPAMMVPPQAQPLVPPLLSPLPALPGPPINPLAMLLALPMTTPA   MSL 727
           :  |  : || :  : :| |      :: | :|    ||
Db    2727  PAPAPTPAPP  LPTPAPPAMPAPA -         -------TPPPTPPAPAPPPPL 2765

Qy     728 VLPVQNPRHQHQAQTSGATGLTTLPAPPAPPLALPPPPAGLLPGVGLPPPTPPPA 785
           :|:      |: |  | :: |             ||| |      |
Db    2766 PLAQPPGPL        LPSPSPAPVPAPGSELVP-----  -   PPSG-GALGSPPSPV 2806
```

```
Qy     776 ESPPGPPPSPQQPPDYPLPMAQGPIPVPTVPPPARVLPLPPMSETPPPPLETAPLLPP 834
           | ||| |                 :::    : :|:::| : :                ||
Db    2807 ---PSPPPSL------------EPVPGLPL PPSRVLAPVP              APAPPPA 2850

Qy     835 -  PPPPPVPPPPLPAAPPPAVPPPPAPPAQPPVTLPPTLPPPPPNSALPPPPPPVL  894
           | ||;  ::       |:        :|     :  :  :      : : :  :
Db    2851 PPPPAPPPAPERPAPPPAPE       RPAPPPAPPPPAPPP       APPPPAPPPAPE 2894

Qy     895  PVQSSNSAPPPFQLPRPPPNII  PMPQSPPPPVAVPVPLPPPP  PPPLPKPN    PP 941
           |  :  |     | ||  :  :   : :  | |   |  | .:        :|'    |
Db    2895 PPAIAVQSPSPPAPSNALVPLVAPALPPPPPPPPPPPAPAAPPPPANIAPPNTPAVPP 2954

Qy     942              PPPALAPPPINPVPH---   PPPPPPNLPPPPP      AVVPP 979
                        :  | : :|:::      ||: :  :  :  :      | :'
Db    2946 TVIQSSPSALSPLPSAPPPAPPPAMPVVAPPPPAPPAAPPAPAPVAPPAPPPAPAAATPA 3005

Qy     980 PPAPPPLEPSLSLPPPVPPPPPPPPPSP- NPPPTAP  APP PPPALPPPLPPPPPPPP 1037
           |  | : | ::          : :|  ||: :|  .  .:::: : :  :  :
Db    3006 PPIAPPIAPPIA---  ---PPIAPPIAPPIAPPIAPPIAPPIAPPIAMPAPPPAAPAPAA 3057

Qy    1038 VESHALTPPMEDPNISTKLLVP 1062
             :  :  |     :  | |
Db    3058 SSAPPIAPPATPAPLIAPLAAP 3079


RESULT 13
Q9Y6V0
ID   Q9Y6V0      PRELIMINARY;      PRT;  5147 AA.
AC   Q9Y6V0; O60305; Q9UIV2; O43473; Q9Y6U9; Q9BV78;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Piccolo protein (Aczonin) (DJ0897G10.1)
DE   (Fragments).
GN   PCLO OR ACZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-759 FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550 kD putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with Rim and Bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   SEQUENCE OF 552-4404 FROM N.A.
RA   Waterston R.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC   TISSUE=BRAIN;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA   Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [4]
RP   SEQUENCE OF 4405-4439 FROM N.A.
RC   TISSUE=PLACENTA;
RA   Strausberg R.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 4625-5147 FROM N.A.
RA   Kalicki J., Elliott G.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: MAY ACT AS A SCAFFOLDING PROTEIN INVOLVED IN THE
CC       ORGANIZATION OF SYNAPTIC ACTIVE ZONES AND IN SYNAPTIC VESICLE
CC       TRAFFICKING (BY SIMILARITY).
CC   -!- SUBUNIT: INTERACTS WITH PAPAC1/PRA1 AND PPCPILLIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CONCENTRATED AT THE PRESYNAPTIC SIDE OF
CC       SYNAPTIC JUNCTIONS (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2,
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: C2 DOMAIN 1 IS INVOLVED IN BINDING CALCIUM AND
CC       PHOSPHOLIPIDS. CALCIUM BINDS WITH LOW AFFINITY BUT WITH HIGH
CC       SPECIFICITY AND INDUCES A LARGE CONFORMATIONAL CHANGE.
CC   -!- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; Y19188; CAB60727.1; -.
DR   EMBL; AC004903; AAD20936.1; -.
DR   EMBL; AC004886; AAD21789.1; -.
DR   EMBL; AB011131; BAA25485.1; -.
DR   EMBL; BC001304; AAH01304.1; -.
DR   EMBL; AC004082; AAB97937.1; -.
DR   HSSP; P04410; 1A25.
DR   MIM; 604918; -.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR002965; P_rich_extensn
DR   Pfam; PF00168; C2; 2.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00239; C2; 2.
DR   PROSITE; PS00479; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN; 2.
KW   Calcium/phospholipid binding; Zinc; Metal-binding; Zinc-finger;
KW   Repeat; Alternative splicing
FT   NON_TER      1      1
FT   DOMAIN     400    465       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                               P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING    499    523       C4-TYPE (POTENTIAL).
FT   ZN_FING    869    892       C4 TYPE (POTENTIAL).
FT   NON_CONS  1010   1011
FT   DOMAIN    2300   2425       POLY-PRO.
FT   DOMAIN    4391   4442       PDZ
FT   DOMAIN    4544   4633       C2 DOMAIN 1.
FT   DOMAIN    5021   5121       C2 DOMAIN 2.
FT   VARSPLIC  4404   4404       S -> SGNGLGIPIVGGKEIPGHSGEIGAYIAKILPGGSAE
FT                               QTGKIMEG (IN ISOFORM 2).
FT   VARSPLIC  4534   4534       K -> KPTPGTKVVSHPITGEIQ (IN ISOFORM 2).
FT   VARSPLIC  4576   4576       G -> GQVMVVQNAS (IN ISOFORM 2).
FT   VARSPLIC  4757   4761       TAHKS -> SKPRK (IN ISOFORM 2).
FT   VARSPLIC  4762   5147       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   5147 AA;  563537 MW;  CD5D84990498CD3C CRC64;

  Query Match             4.7%;  Score 328,  DB 4,  Length 5147,
  Best Local Similarity   21.9%;  Pred. No. 1.1e-11,
  Matches  251;  Conservative  125;  Mismatches  378;  Indels  302,  Gaps   61;

Qy     239 DLGRL--EIPAQLATLLEFAEGHQAILTGSITESLPPEVPAPPSLTLPPDIDQPPFSSFV 296
           || :|   |   |:| :: ||:|   |  ||:     ||      ||      :|    |'
Db       8 DLSQLSPPFPPQIAAVMSPAQG---LPKGSV----PPAAAESPSMHRKQELD----SSHP 56

Qy     297 DTDFQPFEIIPPQLPIDLTFIGENFQLALEINRVMLRLLGEGSLQSWQEQTMGTFLV 356
                 :|  | ||:|    |::       :  :  |   .| |           |  .
Db      57 PKQSGPP  PDPGPPAQPGLSF ---- SPTTDTPPSPQKLPG-     ----RSPSTISL 99

Qy     357 GLALPPDSIPPELPQLIVAQIWPNPIPQLNQRGWALMVILLSSFAPTPALEKPLLKF--- 413
           :::: |  |::|  |           ::|  ||      |:   : ||
Db     100 KPSKSPTDLKEEHKS     ...     ...     SMMPGFLSEVNALSAVSSVVNKFNPF 140

Qy     414 --VSDQAPSGMAALCQHKILGALEQTP--LAPMASPSHPPTQLEWKAGLPPGPMALDVFT 469
             :||    |     : |::   : |   .     .  ||  . . | |  .  |
Db     141 DLISDSFASQFFTTKFQVVVQFFQGFPPGIIKPPLQQQPPKPIPKQQCDGPDPLQQPQTF 200

Qy     470 PNPEESYGAF-VPSWT--TGEQFAG----------WILQSRGLEAPP--   ---RGWSV   507
           :  |   | ::|    ||:   |             :   |    |        || ||
Db     201 PAIQPQQPPKIPKQPPGTQPPIQTPITDHAPLPLQFDASPFQTKQADIVPGPSVKP 260

Qy     508 SDHS----------GDAWPDLPGCDFVLDLDSQTFDLGDPAGPHNYPITPLGLAESIPP 556
           || |           |    ||           :  | ||   |  | ||  :  :
Db     261 SNPSPSKPPIQQPTPGKPPAQQPG----------HPKSQP GPAKPPAQPSGLTKPLAQ 308

Qy     557 APG-VQAPSLPPGLPPGPAPILASSFPPGE---ASKPENLDGFVDHLFEPALAPGFSDLE 612
           || |: |  |||    ||  |  ::|| :   : || :                  ||   |
Db     309 QPGTVKPPVQPPGTIKPPAQPLGPAKPPAQQTGSEKPSSEQ---          --PGPKALA 356

Qy     613 Q--GWALSPKMPGGGSVGITQQ     -----     -----  SYPNVYFSNVQAPGVGPAM 661
           |  |   :   : | :  ||||                        |   |   : || ||
Db     357 QPPGVGKTPAQQPGPAKPPTQQVGTPKPLAQQPGLQSPAPAPGPTKTPAQTKPPSQQPGS 416

Qy     662 I-PAPMPVMTAMGAVPTMPAMMVPPQPQPLVPSLDSPLPALQLQNFINQQAMILAQQMTT 720
           | |    ||  : |  |    ||  ||      | : : ||  :      |  |||  |
Db     417 TKPPPQQPGPAPPS-PQQPGSTPPPSQQP-----GSAPPSAQQPS--  PAPPSAQQPTP 466

Qy     721 QAMSLSLBQQNQPHQHQAQTS-GATSQPPPSTTAPKAKKPPAP---------------- 762
                   :||  |   |||  | :| ||:||:
Db     467 PV--------------SQTGPGKPLQPP--TVSPSAKQPPSQGLPKTIQPLQNITPLLL 509

Qy     763 --QEKPESN-------  ----LEPSGVGLPETTPEPAEPPQPPPSPLQPETVPLPMG 796
             ||   |                |:          |   |:|     : | ||     :|
Db     510 HVPEKANPNTGTPGQTTVQSLGGPNPN  ------PHLTEAPEWLQTNTQMPE    ALG 557

Qy     797 QD--PI------RVKT--VKPPAKVQIPQEEMEETEEEED    -ETAELSPPPPPPPVV 841
           | |  |      ::|| |   : |     ::|  ::|      : ::   |  |||:|
Db     558 GDLAPVPSSPQPKLPTAPVTTTSAVSKSSPQPQQTSPPKPAAPPQPLSKAPEPKKPPPLV 617

Qy     842 PPPLPASPD--PAVPEPEAE----   PALEEVPTQGELPPVPSESNSAPQHPP----    885
           |:|     |  || :  ||:       |  :  ||: :  ||  ::  |: ||
Db     618 PQPTLHGSPSAKAKQPPEADSLSPPAPPKEPSVPSPQPPAPV- ADDPPPQPPMVPPTTP 675

Qy     886 -    -------- PS------     RVPPVQS-----SNSAPPR-----------PQPS 905
                        ||           ||:::    |  | ||            |:|:
Db     676 LVSSSSATTKPDIPSSKVQSQAEEPTTPPLPETDSAKPSQEPPTGERVIPPDSKAIPPPA 735

Qy     906 PEIPNITPMYQSPPGPVAV---  -  PVPTTPPIPIPQPPNIPLPLEALAKLGINGVHLP 957
           : : |      | |||  :       |||      |  |  |  ||  :  ||     |  |
Db     736 SDSKII-----SHPGPSSESKGQKQVPPVQPKEEPPKPAQTPMSPKPD--AKPMPKGSPTP 788

Qy     958 LSTSPNQGKSSP-PAVVPEPPAE-----      ----PPLEPS          ----      985
             |  |.. | |    |:|:                    |: :|:
Db     789 PGPPPTAGQTVPTPQQSPPPQPGPPPSLNLGSITDAPPSQPTTPQETVTGKLPGPGASI 848

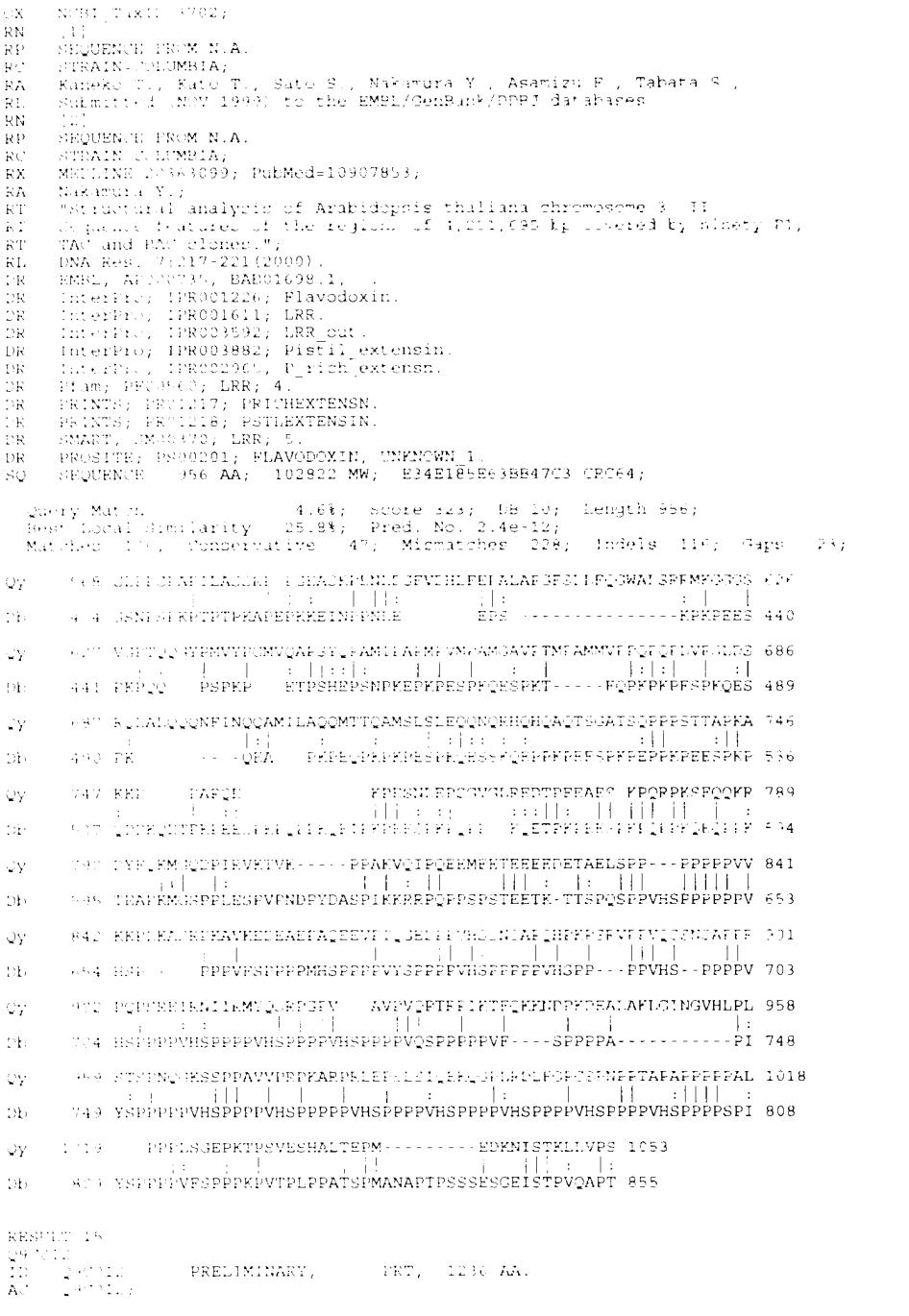
Qy     986 ----   -LSIQEKQGPLRDLPGPCGP     ----NPPTAPAP       --PPPIALPPPLS 1023
                  .|    : ||      || :|       |||: |      |  ||  ||    |:
Db     849 PPQAPNLTSTAQQPSP-HSQSPGAPMPLAPAPSQPPTSQHTPPTTQLAPPAPAPSIPVPL 907

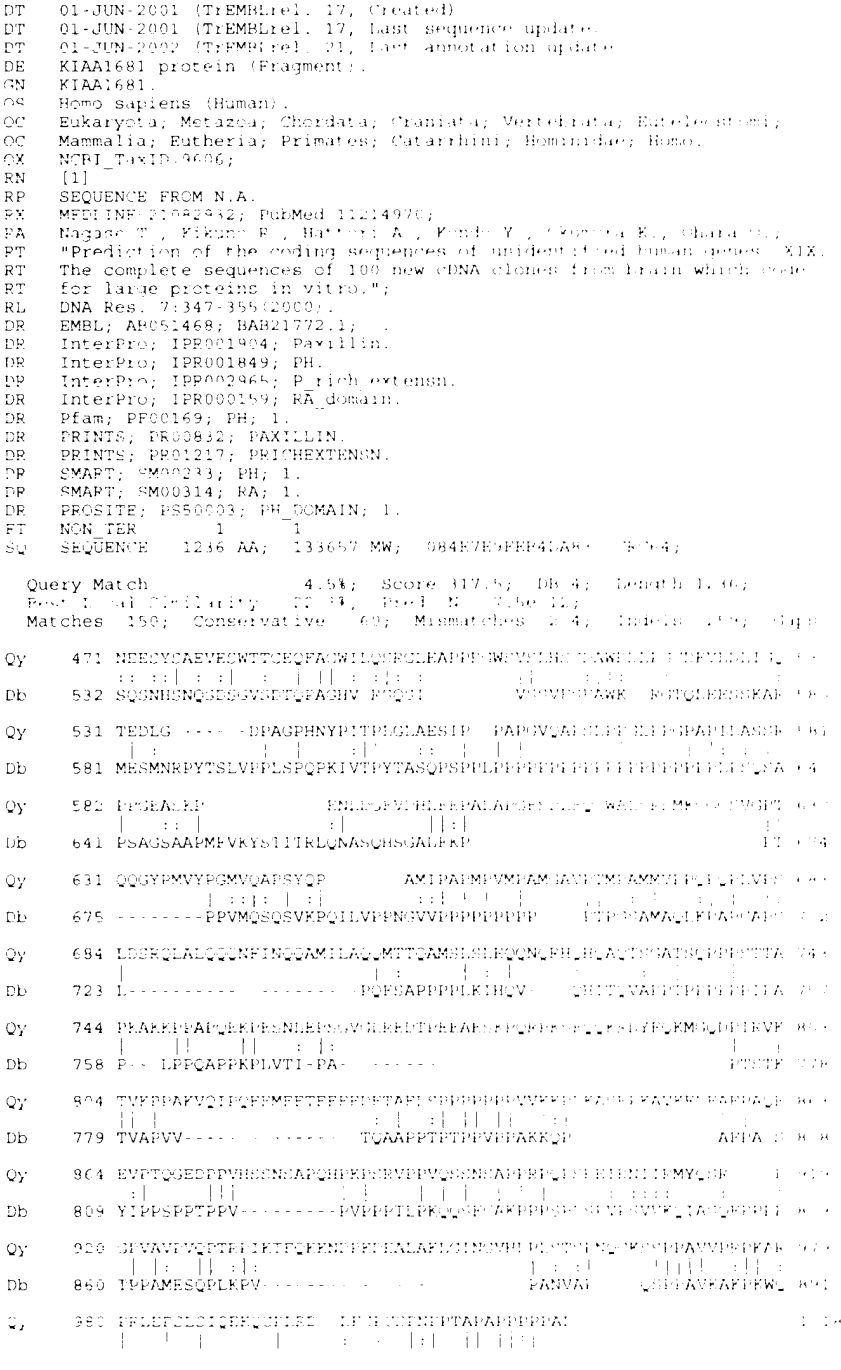
Qy     1024 GPPKTPSVESHALTPPMED-     PNTSTKLLVPGGSVGPSYANAPWPLPL PPEVPTPR 1077
            | | |: |    ||  :  |     |  |:  |      |  |  | | |  |    |:
Db     908 PPTPAPAAEK --LPTPAELAPTVPTTETEPPTPPTPTSPSLTAEPLPAVLPTPLPL PP 964

Qy     1078 PNPSHPYCLSLL------------QQTLRDT-----------PTESCTRISQDLPHPMK 1114
            . | | . |              | :                  ||:|    | :|  :
Db     965 PESTCPLCPTPLNIGSPDPPNPNTPTPPPLQVCNLGSPNPTTPLLIENS  - QTLK AL.. 1019

Qy     1115 GLLGDL 1120
            | |||:
Db     1020 GQLGDI 1025


RESULT 14
Q9LJ64
ID   Q9LJ64      PRELIMINARY;      PRT;   956 AA.
AC   Q9LJ64;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Extensin protein-like.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```


CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20383099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 1,011,695 bp covered by clones
RT TAC and PAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP000735, BAB01698.1, -.
DR InterPro; IPR001226; Flavodoxin.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_cut.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00960; LRR; 4.
DR PRINTS; PR01217; PRICHEXTENSN.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR SMART; SM00370; LRR; 5.
DR PROSITE; PS00201; FLAVODOXIN, UNKNOWN_1.
SQ SEQUENCE 956 AA; 102822 MW; B34E1B5E63BB47C3 CRC64;
Query Match 4.6%; Score 323; DB 10; Length 956;
Best Local Similarity 25.8%; Pred. No. 2.4e-12;
Matches 174; Conservative 47; Mismatches 328; Indels 114; Gaps 24;
RESULT 15
PRELIMINARY, PRT, 1236 AA.


DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1681 protein (Fragment).
GN KIAA1681.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082982; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
DR EMBL; AB051468; BAB21772.1; -.
DR InterPro; IPR001904; Paxillin.
DR InterPro; IPR001849; PH.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR00832; PAXILLIN.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00233; PH; 1.
DR SMART; SM00314; RA; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
FT NON_TER 1 1
SQ SEQUENCE 1236 AA; 133657 MW; 984E7E9FEB4CAB4C CRC64;
Query Match 4.5%; Score 317.5; DB 4; Length 1,236;
Best Local Similarity 25.8%; Pred. No. 7.5e-12;
Matches 150; Conservative 60; Mismatches 274; Indels 139; Gaps 26;

```
Db    892 PSSIPVPSPDFPPPPPESSLVFPPPPPSPVPAPPPPPPPTASPTPDKSGSPGKKTSKTSS 951

Qy   1019      PPPL   -------SG----EPKTPSVES--HALTEPMEDKNISTKLLVP----- 1052
                |||              ||    ||| |||:|     | | |  : | : | |
Db    952 PSPPPPPPTPPPPNSPPIPSSSPAPHPEPKEPSVDSLVSKFTPPAESGSPSKETLPPPAAPP 1011

Qy   1053 -SGSVCFSYANAP 1064
           | :  |  | |
Db   1012 KPGKLNLSSVNLP 1024

Search completed: January 10, 2003, 08:22:30
Job time : 116.696 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: January 10, 2003, 02:19:59 ; Search time 43.8703 Seconds
(without alignments)
896.699 Million cell updates/sec

Title: US-09-803-126-1
Perfect score: 6991
Sequence: 1 MRRLPQPATPQA3TEQNPPT ... WLPFSPLGPAQEVPPTEEVK 1332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422022 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	286.5	4.1	1185	4	US-09-041-886-23	Sequence 23, Appl
2	269	3.8	1135	2	US-08-574-959A-7	Sequence 7, Appli
3	269	3.8	1135	4	US-09-357-014-7	Sequence 7, Appli
4	267.5	3.8	1140	4	US-07-757-022B-104	Sequence 104, App
5	267.5	3.8	1263	4	US-07-757-022B-52	Sequence 52, Appl
6	267.5	3.8	1404	4	US-07-757-022B-2	Sequence 2, Appli
7	267.5	3.8	1404	4	US-07-757-022B-62	Sequence 62, Appl
8	264.5	3.8	1049	4	US-07-757-022B-58	Sequence 58, Appl
9	264.5	3.8	1313	4	US-07-757-022B-140	Sequence 140, App
10	264.5	3.8	1314	4	US-07-757-022B-50	Sequence 50, Appl
11	264.5	3.8	1354	4	US-07-757-022B-48	Sequence 48, Appl
12	264	3.8	941	4	US-07-757-022B-14	Sequence 14, Appl
13	264	3.8	1022	4	US-07-757-022B-84	Sequence 84, Appl
14	264	3.8	1038	4	US-07-757-022B-74	Sequence 74, Appl
15	264	3.8	1270	4	US-07-757-022B-44	Sequence 44, Appl
16	264	3.8	1311	4	US-07-757-022B-42	Sequence 42, Appl
17	264	3.8	1320	4	US-07-757-022B-46	Sequence 46, Appl
18	264	3.8	1320	4	US-07-757-022B-60	Sequence 60, Appl
19	264	3.8	1361	4	US-07-757-022B-40	Sequence 40, Appl
20	257	3.7	905	2	US-08-574-959A-9	Sequence 9, Appli
21	257	3.7	905	4	US-09-357-014-9	Sequence 9, Appli
22	250.5	3.6	1274	4	US-09-095-443-2	Sequence 2, Appli
23	245	3.5	2972	4	US-09-579-181-2	Sequence 2, Appli
24	245	3.5	3118	4	US-09-579-181-1	Sequence 1, Appli
25	244.5	3.5	1099	4	US-09-442-100-2	Sequence 2, Appli
26	236.5	3.4	1187	1	US-08-320-559-28	Sequence 28, Appl
27	236.5	3.4	1187	3	US-08-545-860D-28	Sequence 28, Appl
28	236.5	3.4	1187	5	PCT-US94-04496-28	Sequence 28, Appl
29	236.5	3.4	1210	1	US-08-320-559-26	Sequence 26, Appl
30	236.5	3.4	1210	3	US-08-545-860D-26	Sequence 26, Appl
31	236.5	3.4	1210	5	PCT-US94-04496-26	Sequence 26, Appl
32	216.5	3.1	1004	4	US-08-916-352-2	Sequence 2, Appli
33	216.5	3.1	1142	4	US-09-061-709-2	Sequence 2, Appli
34	216	3.1	2442	4	US-09-514-047A-10	Sequence 10, Appl
35	215.5	3.1	1142	2	US-08-333-118-7	Sequence 7, Appli
36	215.5	3.1	1142	3	US-08-845-928C-7	Sequence 7, Appli
37	212	3.0	2441	1	US-08-194-468-2	Sequence 2, Appli
38	212	3.0	2441	4	US-09-514-047A-8	Sequence 8, Appli
39	211.5	3.0	1384	4	US-08-976-255-11	Sequence 11, Appl
40	206	2.9	2414	1	US-08-227-536-2	Sequence 2, Appli
41	206	2.9	2414	5	PCT-US95-04682-2	Sequence 2, Appli
42	206	2.9	2441	3	US-08-061-739-2	Sequence 2, Appli
43	204.5	2.9	1312	4	US-09-041-886-19	Sequence 19, Appl
44	203.5	2.9	650	6	5202236-13	Patent No. 5202236
45	203	2.9	1125	4	US-09-513-793A-152	Sequence 152, App

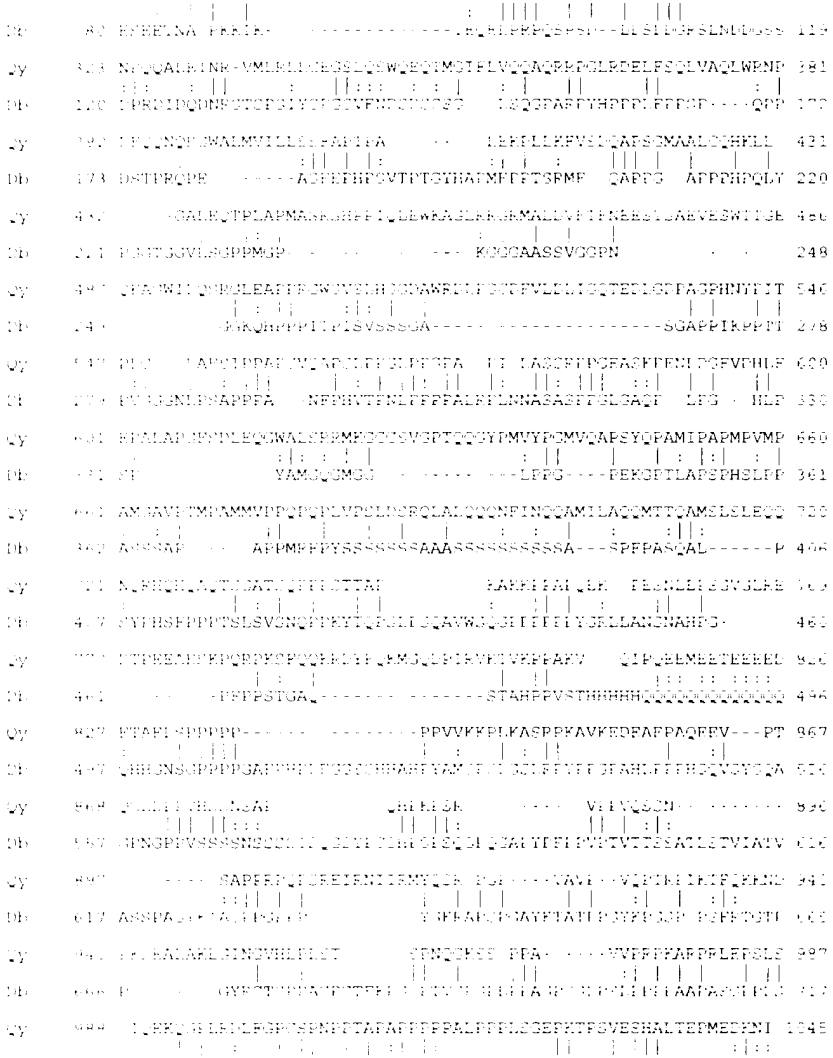
ALIGNMENTS

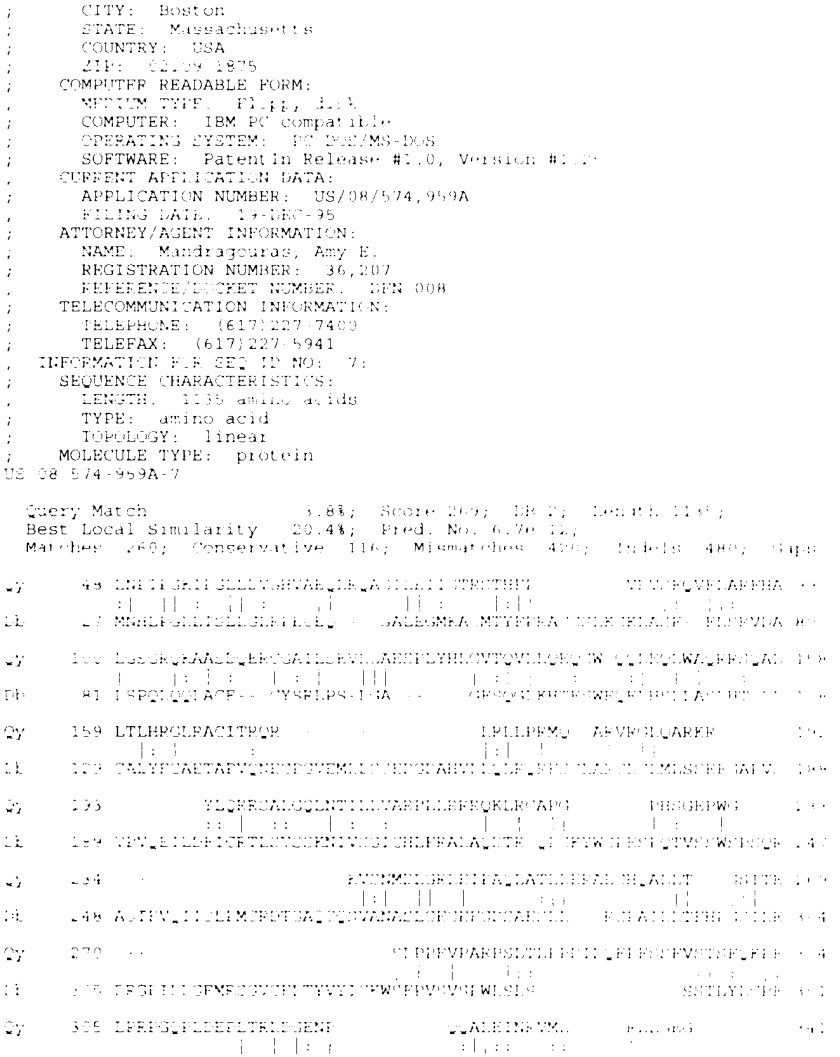
RESULT 1
US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-23

Query Match 4.1%; Score 286.5; DB 4; Length 1185;
Best Local Similarity 22.5%; Pred. No. 3.3e-13;
Matches 216; Conservative 81; Mismatches 313; Indels 349; Gaps 52;

Qy 223 AEGPHS------GEPWGKVSNMDLGPLEIPAQLA--TLLERAESHQALLIG 270
Db 24 APGPEEELPSEGRAEPGPVSTGSSPGPAEPSPQTAPTARVEEACTPPHHKQGPSEEISPS 79

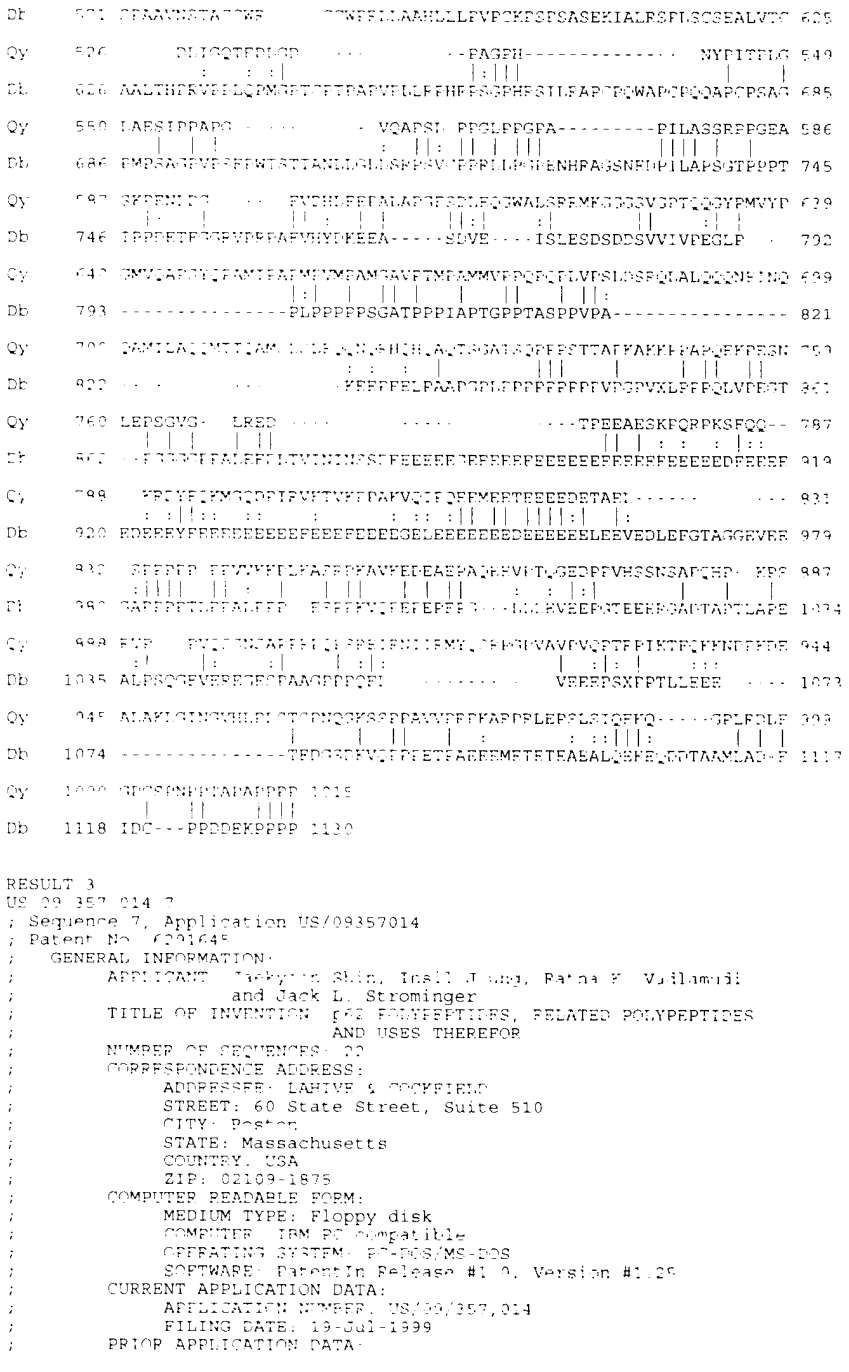
Qy 271 LPPEVPAPPSLTLPPPTPQPPSSFVSTPQKPELPPHGQPLDEPLIRLLG------ E 320





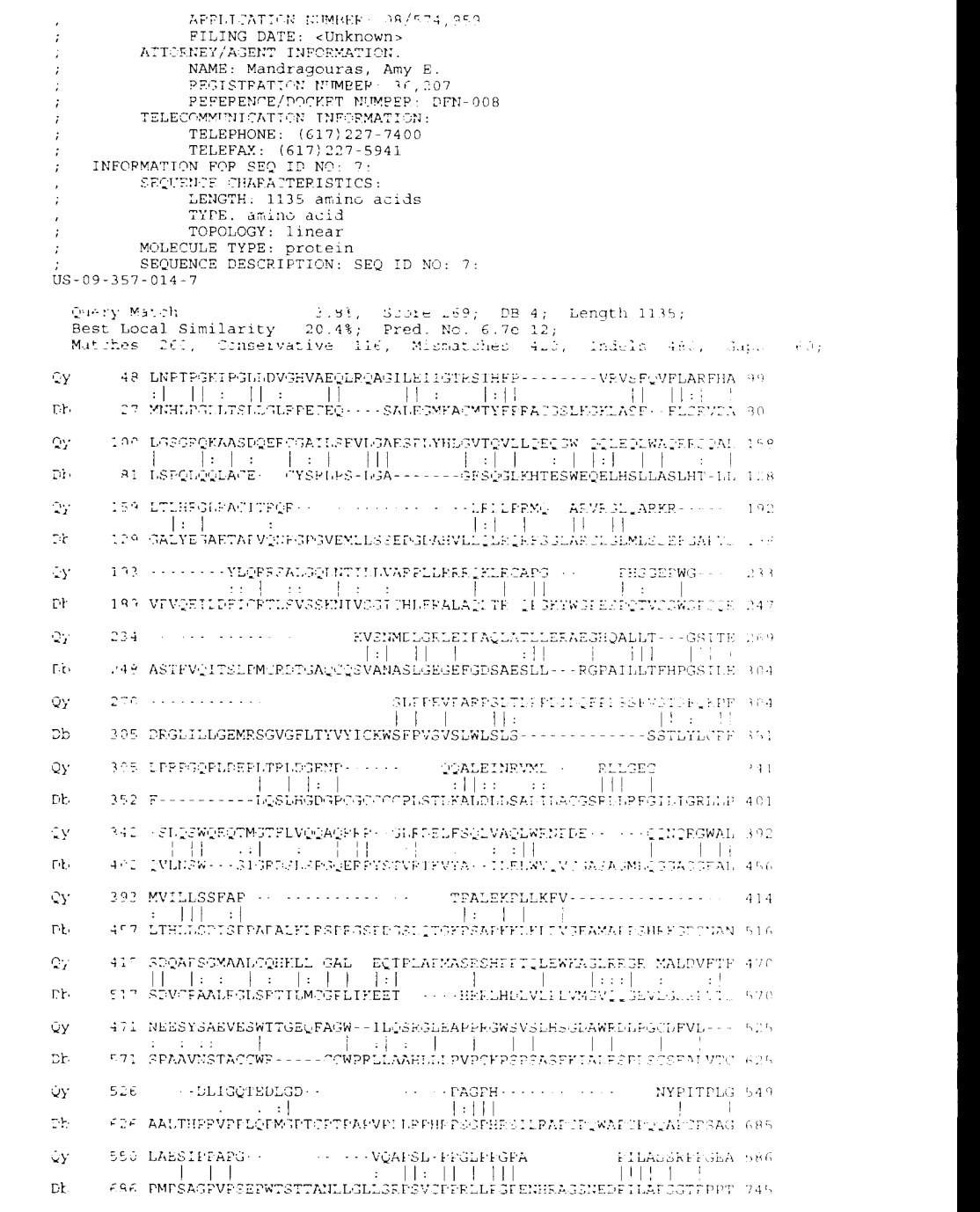
```
RESULT 1
US/08/574,959A/7
; Sequence 7, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
;    APPLICANT:  [illegible] Chin, Insil Jeong, Fatma F. Vallejo, Jr.
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  pGG [illegible], RELATED [illegible]
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN 008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US 08 574-959A-7

  Query Match              5.8%;  Score 209;  DB 7;  Length 1135;
  Best Local Similarity   20.4%;  Pred. No. 6.7e-12;
  Matches  260;  Conservative  116;  Mismatches  420;  Indels  480;  Gaps   47;
```



RESULT 3
US-09-357-014-7
; Sequence 7, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyung Shin, Insil Jung, Rajna F. Vallamidi
; and Jack L. Strominger
; TITLE OF INVENTION: [6] POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,014
; FILING DATE: 19-Jul-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/674,950
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-357-014-7

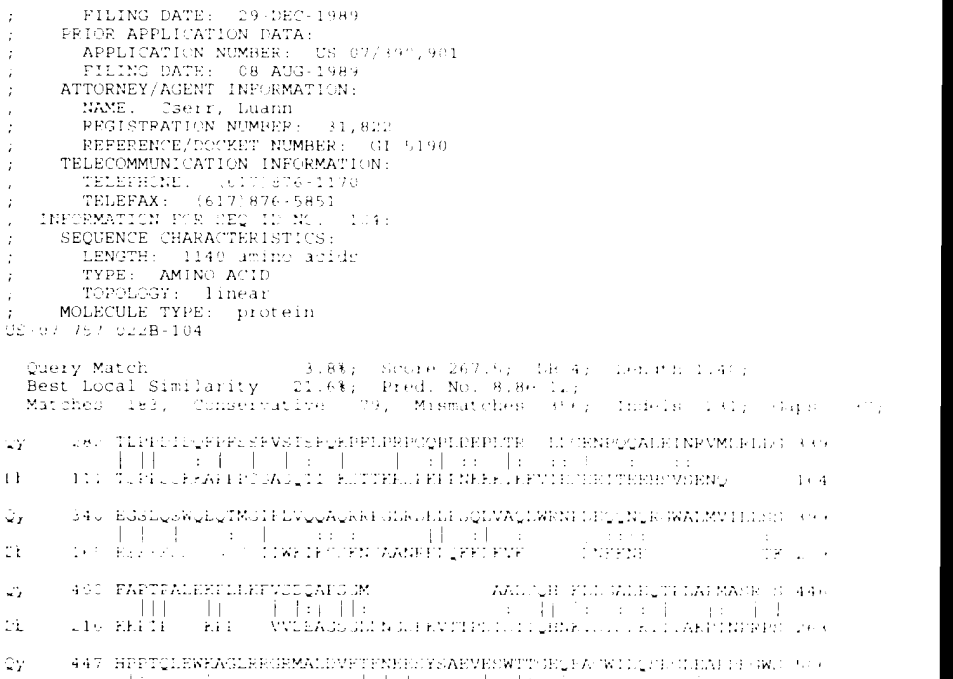
Query Match 3.9%; Score 269; DB 4; Length 1135;
Best Local Similarity 20.4%; Pred. No. 6.7e-12;
Matches 261; Conservative 116; Mismatches 423; Indels 483; Gaps 43;





RESULT 4
US/07/757,022B-104
; Sequence 104, Application US/07757022B
; Patent No. 5433142
; GENERAL INFORMATION:
; APPLICANT: Genetics, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Polypeptides Stimulating Platelets
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridge Park Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 10-SEP-1991
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US/07/757,022B-104

Query Match 3.8%; Score 267.5; DB 6; Length 1,140;
Best Local Similarity 21.6%; Pred. No. 8.8e-12;
Matches 163; Conservative 79; Mismatches 454; Indels 141; Gaps 27;



```
Db     689 APTTPKEPAPT--------TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKE 740

Qy     991 KQGPLRDLFGPCSPNPPTAPAPPPPPALPPPLSGEPKTP----SVESHALTEPMEDKN 1044
                  |   | :|   |||  |  ||    |   | ||   :::  | | |    |
Db     741 PTSTTSDKPAPTTPK-GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP--KK 796

Qy    1045 ISTKLLVPS 1053
            : | | |:
Db     797 PAPKELAPT 805

RESULT 5
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1363 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-52

  Query Match             3.9%;  Score 267.5;  DB 4;  Length 1363;
  Best Local Similarity   21.6%;  Pred. No. 1.1e-11;
  Matches  183;  Conservative   79;  Mismatches  356;  Indels  231;  Gaps   37;

Qy     282 TLPPDIDQFPFSSFVSTSFQKPFLPRPGQPLDEPLTR--LDGENPQQALEINRVMLRLLG 339
             | ||   : |  |  | :  |   |  :| :: |:  :: |    :  ::
Db      70 TSPPSSKKAPPPSGASQTL-KSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQ---  123

Qy     340 ESSLQSWQEQIMSTFLVQQAQPPPGLPPPLPSQLVAQLWENFDEQNQRSWALMVILL.. 399
            | |  |    :  |      :: :         || :|  :      | ::|:        :
Db     124 ESSSSSSSSSSSSTIWKIKSSKNSAANPELQKKLKVK-    DNKKNR-         TK 168

Qy     400 FAPTPALEKPLLKFVSDQAPSGM-----------AALCQH-KLLGALEQTPLAPMASR S 446
             |||   ||    | |:| ||:          :  || |:  : : |    |:  | |
Db     169 KKPTP---KPP-- VVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPS 222

Qy     447 HPPTQLEWKAGLRRGRMALDVFTFNEESYSAEVESWTTGEQFAGWILQSRGLEAPPRGWS 506
             ||   |                  | |:|:     |: || :|         : | |
Db     223 LPPNSDTSKE-----  --TSLTVNKETTVETKETTTTNKQ---  TSTPSKEKTTSAKE 268

Qy     507 VSLHSGDAWRDLPGCDFVL--------DLIGQTEDLGDPAGPHNYPITPLGLAESIPPAPG 559
                   : :||    ||          :   :    |   |   |    | ||   | :  |
Db     269 TQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEP   TETTPKEPASTTPKEPT 326

Qy     560 VQAPSLPPGLPPGPAPILASSPPPGEASKPENLDGFVDHLFEPALAPGPSDLEQGWALSR 619
                 |   |   |||    |  |        :  |:            ||:
Db     327 PTTIKSAPTTPKEPAPTTTKSAP----TTPK-----     EPA     .          356

Qy     620 SMPGGGSVGPTQGGYPMVYPGMVQAPSYQPAMIPAP----  MPVMPAMGA--VPTMPAMM 672
                    ||     |    |     :  |:           |||        |    |     |    |   ||
Db     357  ------  -PTTTKEPA--PTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPT 405

Qy     673 VPPQPQPLVPSLDSRQLALQQQNPINQQAMILAQQMTTQAMSLSLEQQNQRHQHQAQTSG 732
             | :|  |  |                                        |      :  | |
Db     406 TPKEPAPTTPK-- ---------------------------   EPTPTTPKEPAPT-- 429

Qy     733 ATSQPPPST------TAPKAKKPPAPQE-KPESNLEPSGVGLREDTPPEAESKPQPPKSF 785
            | :|  |:|        ||||   |   |:|   |  :   ||:    :|  :|    :  |:  |
Db     430 -TKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSP     TTPKEPAPT 484

Qy     786 QQPPDYFQKMGQDPIRVKTVPPPAFVQIPQEEMEETEEEEDETAELSPPPPPIPVVKKPL 845
             |           |   |    |          |    |  :        |   |   :|   |   |    |||
Db     485 TTK--   ----SAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPA 536

Qy     846 ----KASRPKAVKEDEAEPAQEEVPTQGEDP PVHSSNSAPQHP---------------- 884
                 |   |   ||           ::  ||  ..||       :||   |
Db     537 PTTPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 596

Qy     885 -KPSRVFPVQSSNSAPPRPQPSREIRNIIKMYQSRPGP---VAVPVQPTRPIKTPQKK- 938
            ||:      |  :  :  :   |   | |:                        |  |     |   |   |   |   |   |:
Db     597 EKPAPTTPEELAPTTPEEPTPTT ---         PEEPAPTTPKAAAPNTPKEPAPTTPKEP 647

Qy     939 --NDPKDEALAKLGINGVHLPLSTSPNQGKSSPPAVVPRP--    -KARPRLEPSLSIQE 990
                   ||: |          |   |:|   |  :  |    :   |       |   |:        :  :|
Db     648 APTTPKEPAPT--------TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKE 699

Qy     991 KQGPLRDLFGPCSPNPPTAPAPPPPPALPPPLSGEPKTP        SVESHALTEPMEDKN 1044
                   |    | :|    |||   |    ||    |     | ||       :::   |   |   |    |
Db     700 PTSTTSDKPAPTTPK-GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP    KK 755

Qy    1045 ISTKLLVPS 1053
            : | | |:
Db     756 PAPKELAPT 764

RESULT 6
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
```

```
;       COUNTRY:  U.S.A.
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/757,022B
;       FILING DATE:  19910910
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/643,502
;       FILING DATE:  18 JAN 1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/546,114
;       FILING DATE:  29-JUN-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/457,196
;       FILING DATE:  29 DEC 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/390,901
;       FILING DATE:  08 AUG 1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Cseri, Luann
;       REGISTRATION NUMBER:  31,822
;       REFERENCE/DOCKET NUMBER:  GI 5190
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)876-1170
;       TELEFAX:  (617)876-5851
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1404 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-757-022B-2

  Query Match          3.8%;  Score 267.5;  DB 4;  Length 1404;
  Best Local Similarity   21.8%;  Pred. No. 1.2e-11;
  Matches  184;  Conservative   79;  Mismatches  356;  Indels  231;  Gaps   37;

Qy      282 TLPTTTNQFPPGGPVQTQPQHFLPRPSQFLQRPLTR  LEGENFLQALEIMRVMLRLLS 383
            :  :  : :  :  :  :  :  :  :: ::  :: ::  :  ::
Db      111 TNPPSSKRAIPLSAALQLL PSIPELLPRPGPRFLTFPVIPSPRITPEHCVSENL--- - 164

Qy      384 LSGLQFWQHQTMGTPLVLQQAQRPDLPELEL--QLVAQLWPNFPPLLNLHGWAIMVILSS 441
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      165 LSGLQGWSQCTVGGGGSTLWRLPLLYKISAANPFLVPELPVP     DNPFNP    - -   TP 209

Qy      442 PLLIALRPLLRKVLVPLVAPVQM           AALQQH-FLIQALPQTPIAPMASP-S 496
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      210 KRITP   KTT   TVDEAIGGLSLGNGSFKTTFETITTQNNFVQTIPITTAPTINFSS 263

Qy      497 HPPTQLPWPAGLRPQPMALDVPTPNFFSYGAFVPSWTTGEQFAGWIISSH-LEAPPEGWS 555
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      264 LFFNSVTSKE          TGLTTNPETTVETPTTTTNPG---  TPTTGPEPTTSAPF 309

Qy      556 VSLHSRAWRDLPGTDFVL       -DLIGQTEDLGDPAGPHNYPITPLGLAESIPPAPG 609
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      310 TGSTRRTGAKDLAPTSRVLAKPTPKAETTTPGLALIPKEP- TPTTPFPPASTTPFEPT 367

Qy      610 VGAPSLPPGLPPGPAPILASSRPPGEASKPENLDGFVDHLFEPALAPGFSPDLFQGWAISR 669
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      368 PTTPKLAPTTPKEPAPTTTKEAP    TTPP     ------FPA---------------- 397

Qy      670 PMPSSGSVSPTQQSTPMVTPGMVQAPSTQPAMITAP     MPVMPAMGA- VPTMPAMM 722
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      398          PTTTKEPA  PTTPKEPAPTTTKEPAPTTTKSAPTTPKFPAPTTPKFPAPT 446

Qy      723 VPT-PQPLVPGLPPQLAGQQNPINQLAMILAQMTTQAMSQIPPQMPHQHQAQTPG 780
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      447 TPKEPAPTTPK                          ---------EPTPTTPKEPAPT-- 478

Qy      733 ATSQPPPST------ TAPKAKKPPAPQK KPPSNLPIGVVKLPPITPPAPPFLQLPKPQP 786
            :  :  :  :       :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      471 -TKEPAPTTPKEPAPTAPPPAPTTPPPPAPTTPPPAPTTTKEPTT    TTPPPIAPT 524

Qy      787 QQPRLYPLPMSLQPPIRVPEIVPPIAEVLIPQPPMPPPPPP PIPPLPPPPPVVPPPL 846
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      525 LTP        SAPIIPPPAPIIPSAPIIPKEPGPTIIKEPAPIIKEPAPTTPKEPA 577

Qy      846 ----KAKPPKAVPLLPAPPAQPEVPIQVPIP PVPQQNQAPQM                 889
                :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      578 PTTPKEPAPTTPKEPAPIIKKPAPTAPKEPAPTTPKETAPTTPPIIPTTPEPLAPTTI 634

Qy      890  KPEPVTPVQSCNQAPPPQPPPPIRNIIPMVQVPPQ    VVVVQIPSQIPTPLPP  944
             ||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      638 EKPAPTTPEELAPTTPEPTPTT---      PEEPAPTTPPAAAPNTIKEPAPTTPEP 688

Qy      945   NPEPEALAPLSINSVPPPQTSPINQVPPVPPAVTPPP      PAPPPIPPSLAIQP 999
              ||: :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      689 APIIPKEPAPI---  -   IPKEIAPTTPKPVAPTTPPPAPTTPPPAPPELAPTTPP 740

Qy      999 KQAPLRQLPGPLSPNPPIAPAPPPPAIPPALPPPPPP      PPPPAITPPMPLPP 1053
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      741 PTSIISKPAPIPK-QIAPIPKEPAPIPPPAPPTPPPTAPTTIPPIAPTTP   KP 794

Qy     1054 IQTPIIVPS 1063
            :  :  :  :
Db      797 PAPPPLAPT 805


RESULT 7
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29 JUN 1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cseri, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  62:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-62

  Query Match              3.8%;  Score 267.5;  DB 4;  Length 1404;
  Best Local Similarity   21.6%;  Pred. No. 1.2e-11;
  Matches  183;  Conservative   79;  Mismatches  356;  Indels  231;  Gaps   37;

Qy     390 TLPTTLLQFFRSVPV-IQFLFNQLFRFSIQLIFELTE- LQSENPGQAIFINFVMLFLLS 448
           | |||    : |  | |  :  |  |   :| --  |-  .  |
Db     111 TSPPSSPKPAPPPSGASQTI-KSITKPSPKPPNKKKTKKVIESEEITEEHSVSENQ ---- 164

Qy     449 FSTLSGWIFLIMSTFLVILALFPFSLFFFLFSWLVAQLWRNFFSEQQNQRGWALMVILLSS 508
           | |  |    :  |    :: :      ||  :|  :     | ::|:          :
Db     165 PSSSSSSSSSSSTTWPIPSPPPPAAHPFLIFFLKVF---  TNKFNR-----  ----TK 203

Qy     509 FAPTIALEKFLLKFVSLQAPSGM  ---  ----  AALSQH-KLLGALEQTPLAPMASR S 558
           |||   ||     | |||:          :  || |:  : : |   |:  | ||
Db     210 FFPTP---FPP   VVDEAVSGLDNGDFKVTTFDTSTTQHNKVSTSPKITIAKPINPPPS 263

Qy     559 HPPTQLEWKAGLRRGRMALDVFTFNPECSYSAEVESWTTSEQFAGWILQSRGLEAPPRGWS 618
           ||     |          | |:|:      |: || :|       : | |
Db     264 LPPNSDTSKE-      --  TSLTVNKETTVETKETTTTNKQ- ----TSTDGKEKTTSAKE 309

Qy     619 VSHHSPDAWRQLPSSDPVL------ DLIGQTEDLGDPAGPHNYPITPLGLAESTPPAPG 669
           : :||     ||     :   :   |  |  |   | ||   | : |  |
Db     310 TQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEP--TPTTPKEPASTTPKEPT 367

Qy     670 VLAPSLPPNLPPSPAPILASSPPPGEASPTENLDGFVDNLPEFALAPGFSDLEQGWALSR 729
           |   |  |||     | |    : |:           |||
Db     368 PTTIKSAPTTPKEPAPTTTKSAP----TTPK----------EPA---------------- 397

Qy     730 PMPIPPSVSPTLPVPVVPMVLAPSTGLIAMIPAP-  --MPVMPAMGA--VPTMPAMM 787
                 ||     |   |   : |:        |||      |   |   |  |||
Db     398 ---------PTITKEPA--PTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPT 446

Qy     788 VPPGLLGVVPLPDPGIALLLMPIMIQAMILAQLMTTQAMSLSLEQQNQPHQHLALTSS 845
           | :|| |  |          |                                |  : | |
Db     447 TPKEPAPTTPK-------------------------------EPTPTTPKEPAPT-- 470

Qy     846 ATSQPPLPT-     --TAPPAPPPPAPLPE PPPSNLPPSGVSLPEDTPPEAESKPQPPPKSP 903
           | :| |:|        |||    |  |:|  |  :  ||:     :| :|     : |: |
Db     471 -TKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSP-   TTPKEPAPT 525

Qy     904 QPPKVPLAMNLSPVPETVHPPAKVQIPQEEMEETEEEEDETAELSPPPPPPPVVKKPL 963
           |     |    |    |  |:      | |  :|  | |    |||
Db     526 TTK--------SAPTTPKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPA 577

Qy     964 ----KASPPPAVPPPPAEPAQLPVPTQGEDP PVHSSNSAPQHP------- ------ - 1004
               |    |    ||        ::  ||  ::| |      :|| |
Db     578 PTTPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 637

Qy     1005 PPSPVPFVQSPNSAFPPLQNPFPNTIPMTQPPSP-- -VAVPVQPTPPIKTPQKY- 1058
            ||:   | : : : |  |:          | |     | |  |  |  |  |:
Db     638 EKPAPTTPPFLAPTTPPPPTPTT-----    -PPEPAPTTPKAAAPNTPKEPAPTTPPEP 688

Qy     1059   NQPKPPAIAPLSIINTVPHLPLSTSPINQSPSSLPAVVPPP-- --PAPPPLPPPLPLPLE 1116
              ||: |          |  |:|   | : |  :  |       |  |:     : :|
Db     689 APTTPKEPAPT--------TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKE 740

Qy     1117 PQSPLPDLPGPGSPNPPIAPAPPPPPALPPPLSGEPKTP------SVESHALTEPMEDIN 1170
            |     | | :|   |||  |  |||  |     |  ||       :::  | | |    |
Db     741 PTSTTSDKPAPTTPK-GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP   PK 796

Qy     1171 ISTKLLVES 1193
            : | | |:

Db     797 PAPKELAPT 805

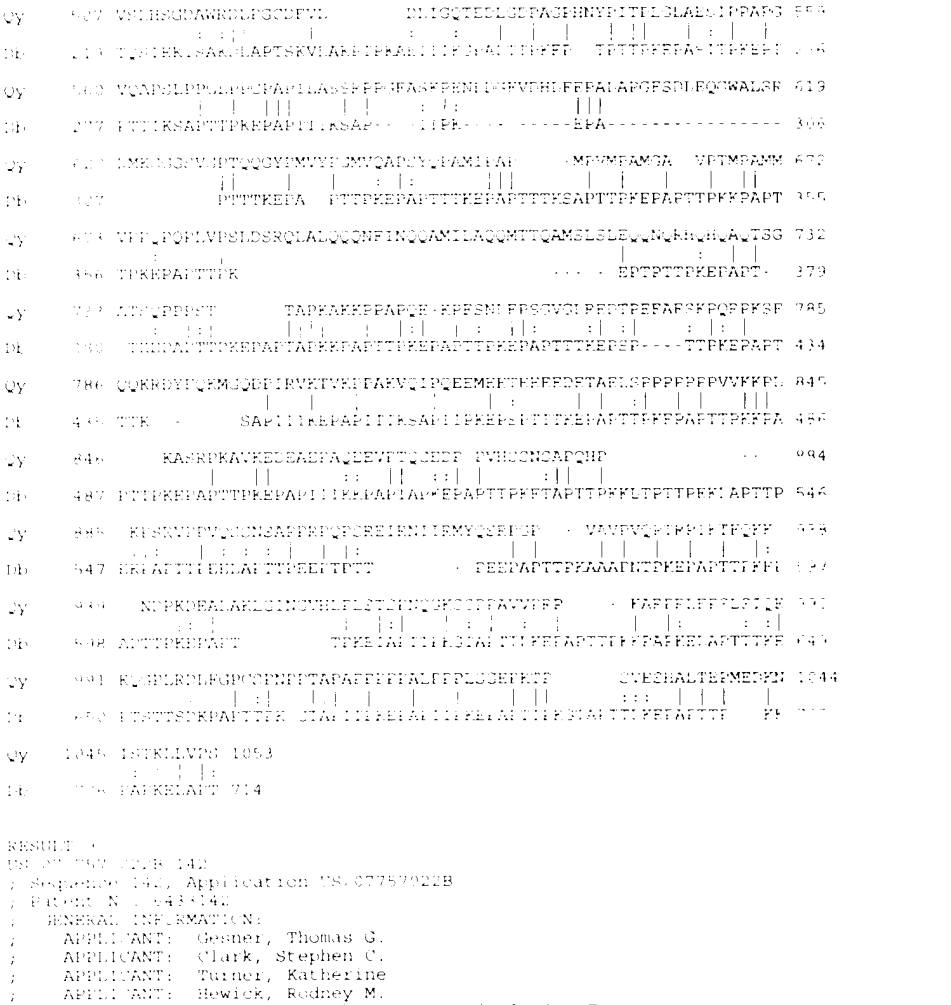

RESULT 8
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1049 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-58

  Query Match              3.8%;  Score 264.5;  DB 4;  Length 1049;
  Best Local Similarity   21.8%;  Pred. No. 1.3e-11;
  Matches  172;  Conservative   68;  Mismatches  326;  Indels  223;  Gaps   34;

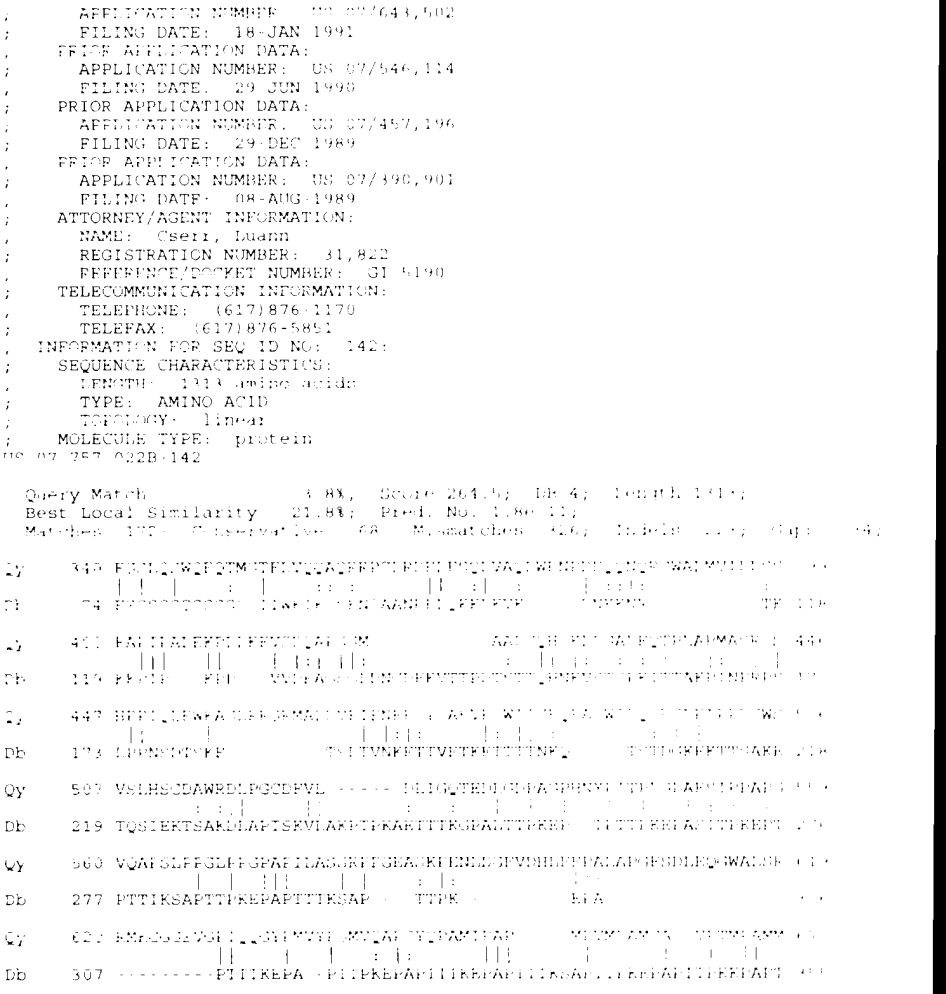
Qy     340 EPSLQSWIEQTMGTFLVILAQFPPSLPFFLFSWLVAQLWRNFFSEQQNQRGWALMVILLSS 399
           | |  |    :  |    :: :      ||  :|  :     | ::|:          :
Db      74 PSSSSSSPPSSSSSTIWPIPSSPPNSAANPFLIKFLKVK ----PNKPNR----------TK 118

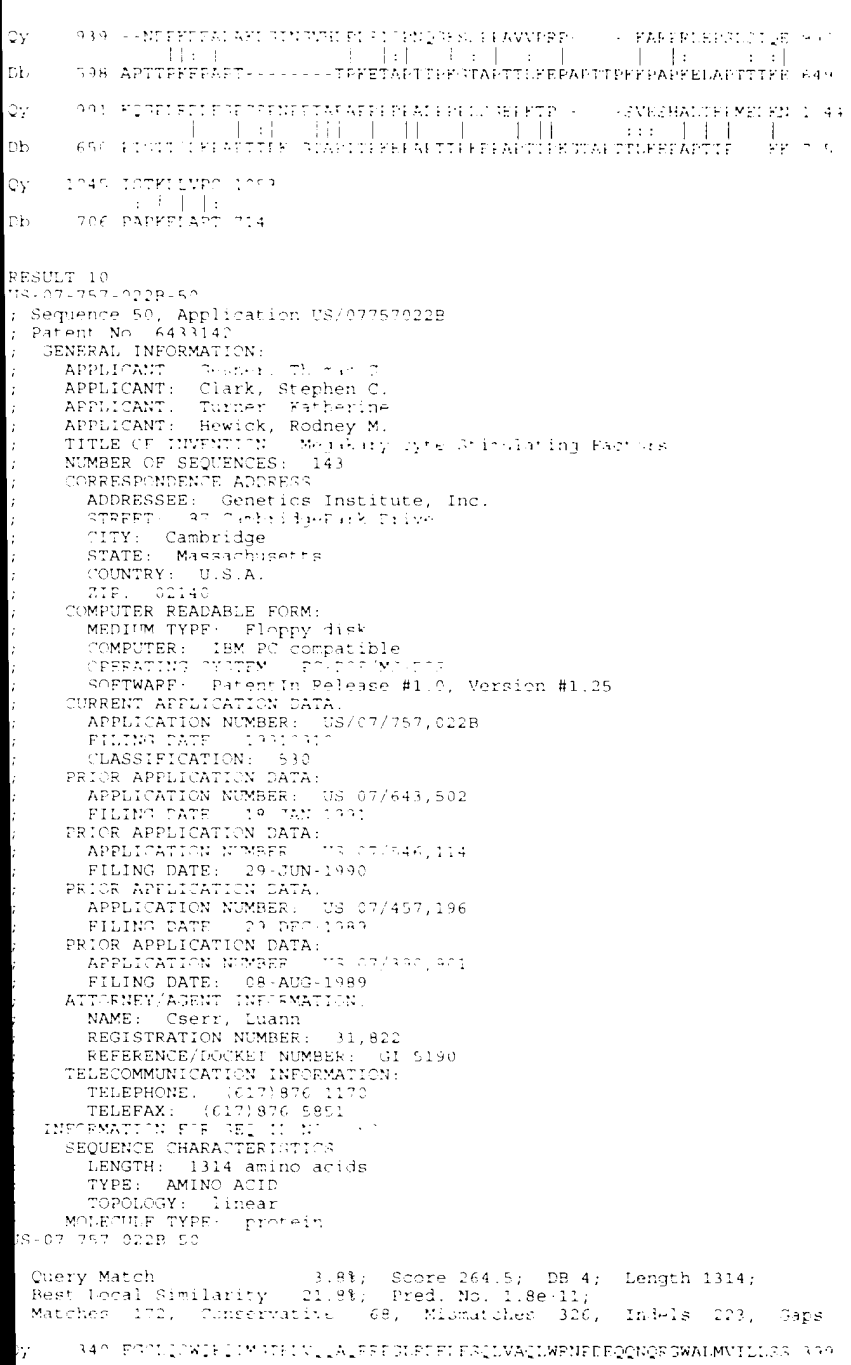
Qy     400 FAPTIALEKFLLKFVSLQAPSGM     ---      AALSQH-KLLGALEQTPLAPMASR-S 446
           |||   ||     | |:|:|:          :  || |:  : : |   |:  | ||
Db     119 KKPTP --KPP --VVDEAGSGLDNGDFKVTTFDTSTTQHNKVSTSPKITIAKPINPPPS 172

Qy     447 HPPTQLEWKAGLRRGRMALDVFTFNPECSYSAEVESWTTSEQFAGWILQSRGLEAPPRGWS 506
           ||     |          | |:|:      |: || :|       : | |
Db     173 LPPNSDTSKE---------TSLTVNKETTVETKETTTTNKQ-     TSTDGKEKTTSAKE 216
```



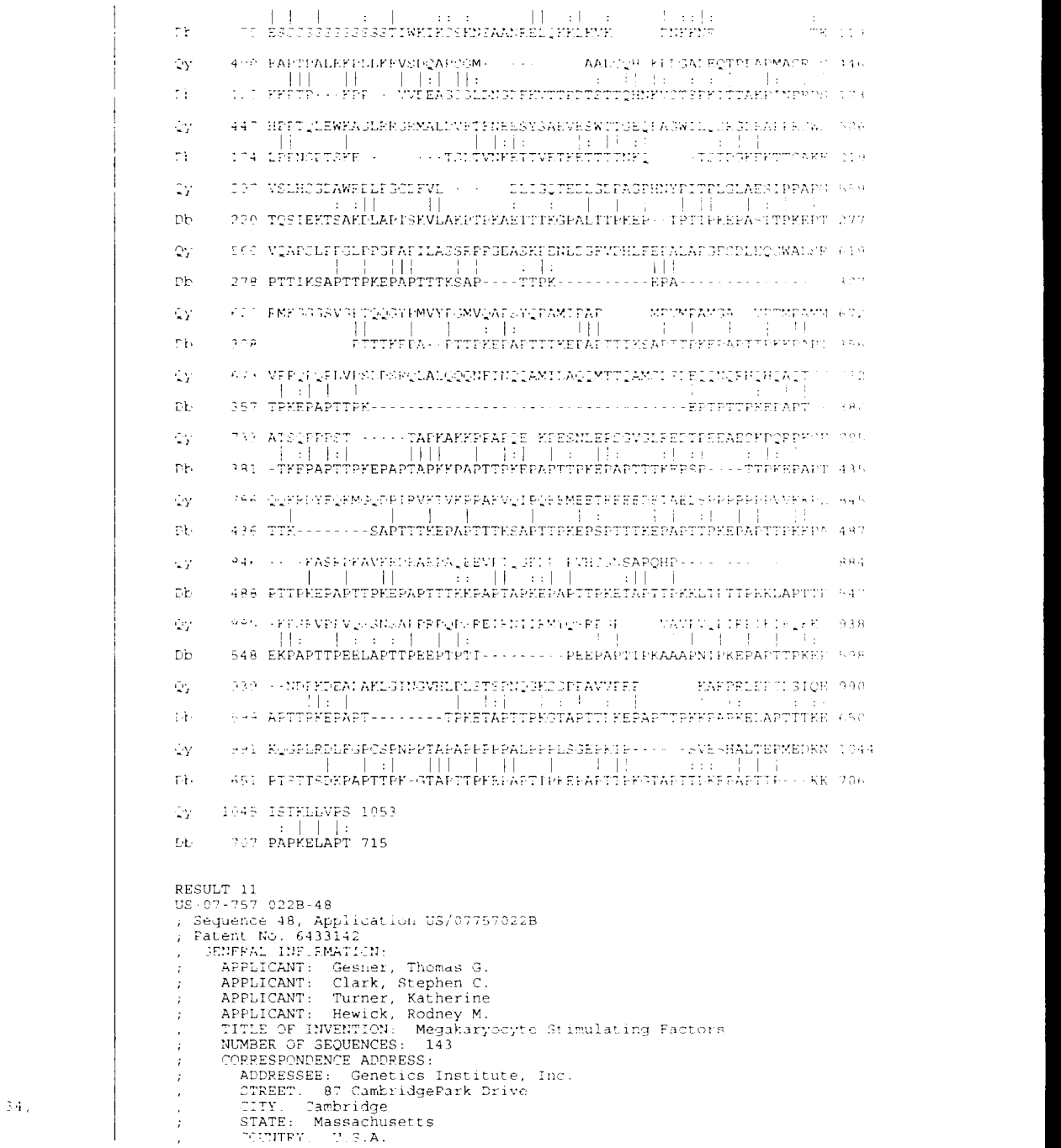
RESULT 1
US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 5443142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Bewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA.
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29 JUN 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1333 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-142





RESULT 10
US-07-757-022B-50
; Sequence 50, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: [illegible]
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 [illegible]
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: [illegible]
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: [illegible]
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 19-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 9190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876 1170
; TELEFAX: (617)876 5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 3.8%; Score 264.5; DB 4; Length 1314;
Best Local Similarity 21.9%; Pred. No. 1.8e-11;
Matches 172; Conservative 68; Mismatches 320; Indels 223; Gaps 34;



RESULT 11
US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.

; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Conry, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876 1170
; TELEFAX: (617)876 9851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-[illegible]-48

Query Match 3.8%; Score 264.5; DB 4; Length 1354;
Best Local Similarity 21.8%; Pred. No. 1.9e-11;
Matches 170; Conservative 68; Mismatches 326; Indels 223; Gaps 34;

Qy 147 EVQLQQWGEGTMSTFLVCQAEEPSLRDSLFDQLVAGLWSCPFPQGNGRCWALMVTLLSS 326
Db 11 ESSSNNSNSSTAATTWAPSSPNSAWSNLLAPELVK 46

Qy 327 PAPTPALEEPLLEPVDLPAPSQM -- -AALDQH-FLLSALFQTPLAPMASP-S 446
Db 47 PRPTP PPP TTTEASSSLDNCPPSTTTPTTSTTQNNPVSTGPPITTAPPINPPI 213

Qy 447 SGTTQLEWKASLRRDRMALDVFTTNEEGYQAEVEQWTTGPQFAGWLLSPCLEAPPPSWS 506
Db 214 LPPNSTTPAA TGLLVNREITVETPPTTTINPQ - - ITTMPPPTTSAPE 259

Qy 507 VSLHSNAWRDLPGSPFVL-- - -DLIGQTEDAGDPAGPHNYPITPLGLAESIPPAPG 559
Db 260 TQSIEETGAPLAPTJPTLAPLTPPAPTTTPTPAPTTPPEP TTTTPPEPAPTTPPEPT 317

Qy 560 TLAPSLPPSLPPGPAPILASSRPPJEASKPPNIDGPVDHLPPPAIAPGPSDLPQGWAISP 619
Db 318 PTTIKSAPTTPKEPAPTTPKEAP TTPP---------EPA------- ---- - - 347

Qy 620 PMKSIRVGPTLQGHMVTDMVLAESYLAMIPAP MEVMPAMGA--VETMPAMM 670
Db 348 TTTTPEPA PTTPPEPAPTTTPPEPAPTTPPSAPTTPPEPAPTTPPEPAPT 398

Qy 671 TPPQLPLVDSLDSRQLALLQQLNPINQGAMTIAIGMTLAMLGLRLWCAVTC 722
Db 399 TKELAPTTPK -------------EPTPTTPKEPAPT-- 420

Qy 723 ATSQIPPSI IAPPARPPAPGR EPESNLSPSGVSLPELTTPEEAPSPPQPPPSP 780
Db 421 TKELAPTTPKEPAPIAPKEPAPTTPKEPAPTTPKEPAPTTTKEPSP----TTPEPPAPT 475

Qy 786 QQKRDYPQKMGQDPIPVKTVKPPAKVQIPQEPMPETPEEEETAPLSPPPPPPVVPPIC 844
Db 476 TTP---- -- SAPTTPEPAPTTPKEAPTTPKEAPTTPKEAPTTPEPAPTTPPEPA 525

Qy 845 PACEPEAVKEDEAPTAQHEVPTTQHPP PVHSNSAPLHI 884
Db 526 PTTPKEPAPTTPKEPAPTTPEPAPTAPKEPAPTTPEPTAPTTPPTTPPEPAPTTP 582

Qy 885 KLKKTIVQLNLAFISLQLSEIRNIIRMCLPS P VSNTVLITTPPTTLVP 934
Db 583 EKPAPTTPEELAPTTPEEPTPTT -- -PEEPAPTTPKAAAPNTPKEPAPTTPKEP 630

Qy 935 --NLTKLEALAKLSINGVHLPLATGPNLHKSSTTAVTPAT PAPPLEPGLNIQE 990
Db 631 APTTPKEPAPI - --- TPEPTAPTTPKGTAPTTPEPAPTTPEPAPTPEPAPTTPEP 680

Qy 991 EQQPLELDLFSPSPNPPTAPAPPPPALPPPLPPPTPT VPVHAITEPMMPLPN 1044
Db 681 PTSTLAEPAPTTPK GTAPTTPEPAPTTPEPAPTTPK PAPTTPEPAPTTP PP 730

Qy 1045 ISTPLIVPC 1053
Db 747 PAPKELAPT 755

RESULT 12
US-07-757-022B-14
; Sequence 14, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Conry, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876 1170
; TELEFAX: (617)876 9851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```
;      LENGTH:  941 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-14

  Query Match              3.8%;  Score 264;  DB 4;  Length 941;
  Best Local Similarity   22.4%;  Pred. No. 1.2e-11;
  Matches  163;  Conservative   58;  Mismatches  208;  Indels  208;  Gaps   32;

Qy     402 PTPALEKPLLKFVGDQAPSGM----      -AALQQH KLLGALEQTPLAPMAGR SHP 448
           |||   ||    | |:| ||:              . || |.  . : |   |. | | |
Db      13 PTP---KPP---VVDPAGSGLDNGDPFVTTPDTSTTQHNVVSTSPKITTAKPINPPPSLP 66

Qy     449 PTQLEWKAGLPPGPMALDVPTPNEESYGAFVERWTTSPLFAGWILQGPGLEAPPPGWSVS 508
           |     |           | |:|:     |: || :|       : | |
Db      67 PNGDTSKP   ...   TSITVNPKFTTVSTPTTTTNFQ ----TSTDGPEKTTSAKETQ 112

Qy     509 IHGGDAWPDLPGGPFVL --     PLLSQTELLGDFAGPHNYDITPLSLAESIPPAPGVQ 561
           : :||      ||      :   :  | |  |  | ||    | : |  |
Db     113 SIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEP- TPTTPKEPASTTPKEPTPT 170

Qy     562 APSLPPGLPDGPAPILASSPPPGPASPFENLMGFVDHLFEPALAPGPSDLEQGWALSRRM 621
           |   |  |||    | |    : |:          |||
Db     171 TIKSAPTTPKEPAPTTTKSAP----TTPK ----      EPA--   -------   --- 198

Qy     622 PGGGSVGPTQQGYPMVYFSMVQAPSYQPAMIPAP---- MPVMPAMGA  VPTMPAMMVP 674
                 ||    |   | : |:          |||      | |   |   |  ||   |
Db     199 -------PTTTKEPA--PTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTP 249

Qy     675 PQPQFLVPSLDQPQLAIQQQNFINQQAMILAQMTTQAMSLSLEQQNQAHQHQAQTSGAT 734
           :| | |                                    |     | | |
Db     250 KEPAPTTPK--------------------------------EPTPTTPKEPAPT-- T 272

Qy     735 SQPPPST-  - TAPFAKPPPAPQE PPPSNLEPSGVGLREPTPEEAESKPQRLKSFQQ 787
           :| |:|      |||| |  |:|  | .  ||:    :| :|    : |: |
Db     273 KEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSP----TTPKEPAPTTT 328

Qy     788 PPPVFQPMMQLPPFVPTVPFIAPVLPQEEMEETPPEDETAELSPPPPPPVVKKPL  845
           |     |   |   |     |   | :    | |  :|  | |   |||
Db     329 P---------SAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPT 380

Qy     846 --KASPPKAVKPDPAEPAQPPVPTQGEDP-PVHSSNSAPQHP -------- -- ----P 885
             |   |   ||      ::  ||  ::| |      :||  |                |
Db     381 TPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPKKLAPTTPKK 440

Qy     886 PQPVPPVQPPNSAPDPPLPSTSPPNQGKSSPPAVVPPP  --VAVPVQPTPPIKTPQPK--- 938
           |:   | : : :  |  | |:              | |    | |  |  |  | |:
Db     441 PAPTTPEELAPTTPEEPTPTT---------PEEPAPTTPKAAAPNTPKEPAPTTPKEPAP 491

Qy     939 NDPKPEALAKLSINGVHLPLSTSPNQGKSSPPAVVPPP  ----PAPPPLPPSLSIQPPQ 992
           ||: |             | |:|   | : |  :  |       | |:      : :|
Db     492 TTPKEPAPT-----    TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPT 542

Qy     993 GPLPDLPPQGPNPPTAPAPPPPPALPPDLSGPPPTP ---- SVPSHALTEPMEDPNIS 1046
           |   | :|   |||  |  ||    |  | ||      :::  | | |   | :
Db     543 STTSPKPAPTTPP-GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP   KPPA 599

Qy    1047 TKLLVPS 1053
           | | |:
Db     600 PKPLAPT 606


RESULT 13
US-07-757-022B-84
; Sequence 84, Application US/07757022B
; Patent No. 6433142
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,332
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,922
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  84:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1022 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-84

  Query Match              3.8%;  Score 264;  DB 4;  Length 1022;
  Best Local Similarity   22.4%;  Pred. No. 1.4e-11;
  Matches  163;  Conservative   58;  Mismatches  208;  Indels  208;  Gaps   32;

Qy     402 PTPALEKPLLKFVGDQAPSGM----      --AALQQH-KLLGALEQTPLAPMAGR SHP 448
           |||   ||    | |:| ||:              :  || |:  : : |   |: | | |
Db      94 PTP---KPP --VVDPAGSGLDNGDPFVTTPDTSTTQHNVVSTSPKITTAKPINPPPSLP 147

Qy     449 PTQLEWKAGLPPGPMALDVPTPNEESYGAFVERWTTSPLFAGWILQGPGLEAPPPGWSVS 508
           |     |           | |:|:     |: || :|       : | |
Db     148 PNGDTSKP------   TSITVNPKFTTVSTPTTTTNFQ--    TSTDGPEKTTSAKETQ 193

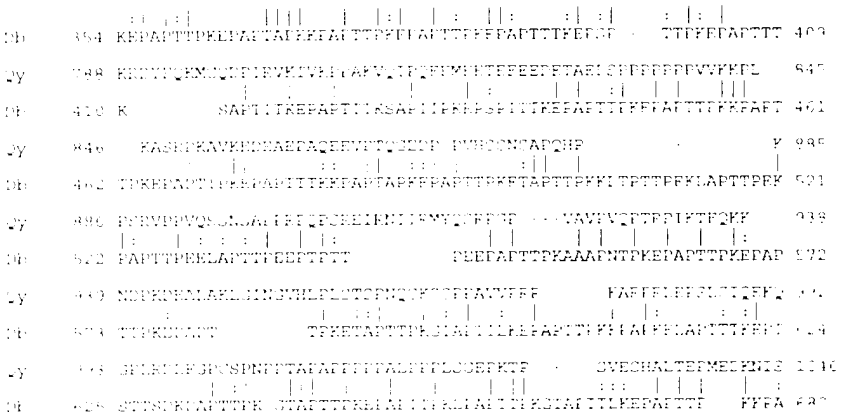
Qy     509 IHGGDAWPDLPGGPFVL      --PLLSQTELLGDFAGPHNYDITPLSLAESIPPAPGVQ 561
           : :||      ||      :   :  | |  |  | ||    | : |  |
Db     194 SIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEP- TPTTPKEPASTTPKEPTPT 251

Qy     562 APSLPPGLPDGPAPILASSPPPGPASPFENLMGFVDHLFEPALAPGPSDLEQGWALSRRM 621
           |   |  |||    | |    : |:          |||
Db     252 TIKSAPTTPKEPAPTTTKSAP----TTPK----------EPA---------          279

Qy     622 PGGGSVGPTQQGYPMVYFSMVQAPSYQPAMIPAP     -MPVMPAMGA  VPTMPAMMVP 674
                 ||    |   | : |:          |||      | |   |   |  ||   |
Db     280 -------PTTTKEPA--PTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTP 330

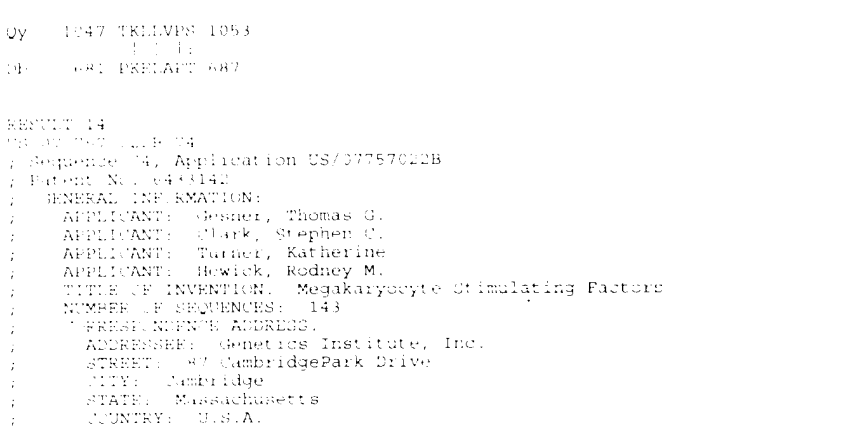
Qy     675 PQPQFLVPSLDQPQLAIQQQNFINQQAMILAQMTTQAMSLSLEQQNQAHQHQAQTSGAT 734
           :| | |                                    |     | | |
Db     331 KEPAPTTPK----------------------------      -EPTPTTPKEPAPT-  T 353

Qy     735 SQPPPST-  -   TAPFAKPPPAPQE PPPSNLEPSGVGLREPTPEEAESKPQRLKSFQQ 787
```



RESULT 14
US-08-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Howick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29 JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Oser, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-74

Query Match 3.8%; Score 264; DB 4; Length 1038;
Best Local Similarity 22.4%; Pred. No. 1.6e-11;
Matches 162; Conservative 58; Mismatches 308; Indels 308; Gaps 5;



RESULT 15
US-07-757-022B-44
; Sequence 44, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.

```
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18 JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1270 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-44

  Query Match             3.8%;  Score 264;  DB 4;  Length 1270;
  Best Local Similarity   22.4%;  Pred. No. 1.9e-11;
  Matches  163;  Conservative   59;  Mismatches  298,  Indels  209;  Gaps   32;

Qy     402 FTPALEKPDLLEFVSDQAPSGM-----------AALCQH-KLLGALEQTPLAPMASR-SHP 448
           |||   ||     | |:| ||:            :  || |:  : :|   |:  | | |
Db      78 PTP---FPP   VVDPAGPGLDNGDFFVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLP 131

Qy     449 PTQLEWKAGLRRGRMALDVFTFNEESYSAEVESWTTGEQFAGWILQSPGLEAPPPGWSVS 508
           |     |             | |:|:      |: || :|       : | |
Db     132 PNSDTGFE-       --TNLTVNFETTVFTKETTTTNKQ    --TSTDGKEKTTSAKETQ 177

Qy     509 LHSGDAWRDLPGCDFVL-------DLIGQTEDLGDPAGPHNYPITPLGLAESIPPAPGVQ 561
           : :||    ||       :   :   |  |  |   | ||   | : |  |
Db     178 SIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEP -TPTTPKEPASTTPKEPTPT 235

Qy     562 APSLPPGLPPGPAPILASSRPPGEASKPENLDGFVDHLFEPALAPGPSDLEQGWALSRRM 621
           |  |  |||    | |    : |:          |||
Db     236 TIKSAPTTPKEPAPTTTKSAP----TTPK----------EPA-------------- --- 263

Qy     622 PGGSSVSPTQQGYPMVYSGMVQAPSYQPAMIPAP     -MFVMPAMGA--VPTMPAMMVP 674
                 ||    |   |   : |:       |||      |  |   |   |  ||    |
Db     264 -------PTTTKEPA--PTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTP 314

Qy     675 PIPIPIVPSSSRLLALQQNFINQQAMILAQLMTTQAMSLSLEQQNQRHQHQAQTSGAT 734
           :| |  |                                   |    : | |  |
```

```
Db     315 KEPAPTTPK----------------------------------EPTPTTPKEPAPT   367

Qy     735 SQPPPST------TAPKAKKPPAPQE-KPPSNLPPSGVGLPPDIPEEMESRPQRPRSPQQ 787
           :| |:|      ||||   |  |:|  | :  ||:    :| :|    : |: |
Db     368 KEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTITKEPSP    TTPKEPAPTTT 423

Qy     788 PPTVPQPVMGQDPTFVPTVPTPAKVQIPQEEMEETEEEELETAELSPPPTTTPVVRKTL  845
           |      |   |   |      |    | :    | |  :|  | | |||
Db     424 P--------SAPTTTKEPAPTTTPSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPT 475

Qy     846   KASPTPAVKEDEAEPAQEEVPTQGEDP-PVHSSNSAEQHP- - - -            K 885
             |  |   ||       ::  ||  ::| |     :||  |                  |
Db     476 TPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKPTAPTTPPPKLTPTTPPPKLAPTTPEK 535

Qy     886 PSPVPPVQSSNSAPPPPQPSPEIPNIIPMYQSPPGP----VAVPVQPIPPIKTPQKE    938
           |:   |     : |  | |.              | |     | |  |  |  |  |:
Db     536 PAPTTPEELAPTTPKEPTPTT---------PEEPAPTTPKAAAPNTPKEPAPTTPKEPAP 586

Qy     939 NDPKDEALAKLGINGVHLPLSTSPNQGKSSPPAVVPPP----  PAPPPLEPPSLSIQEPQ 992
           ||: |             | ||:|   | :|  :  |       | |  |:     : :|
Db     587 TTPKEPAPT--------TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPT 638

Qy     993 GPLPDLFGPGSPNPPTAPAPPPPPALPPPLSGEPKTP      SVESHALIEPMEDPNIS 1046
           |    | :|   |||  |  ||   |    | ||      :::  | | |   | :
Db     639 STTSDKPAPTTPK-GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP---KKPA 694

Qy    1047 TKLLVPS 1053
           | | |.
Db     695 PKELAPT 701

Search completed: January 10, 2003, 02:25:35
Job time : 57.8703 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd

UM protein - protein search, using sw model

Run on: January 10, 2003, 09:03:04 ; Search time 103.409 Seconds
(without alignments)
250.842 Million cell updates/sec

Title: US-09-803-126-1
Perfect score: 6991
Sequence: 1 M[illegible] 1337

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 104601557 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6991	100.0	1337	10	US-09-803-126-1	Sequence 1, Appli
2	2131	30.5	407	10	US-09-803-126-12	Sequence 12, Appl
3	1543.5	22.1	437	10	US-09-803-126-4	Sequence 4, Appli
4	1537	22.0	406	10	US-09-803-126-13	Sequence 13, Appl
5	1537	22.0	406	10	US-09-803-126-6	Sequence 6, Appli
6	1203.5	17.2	333	9	US-09-764-868-758	Sequence 758, App
7	267.5	3.8	1140	12	US-10-124-557-104	Sequence 104, App
8	267.5	3.8	1363	12	US-10-124-557-52	Sequence 52, Appl
9	267.5	3.8	1364	12	US-10-124-557-2	Sequence 2, Appli
10	267.5	3.8	1401	12	US-10-124-557-62	Sequence 62, Appl
11	267	3.8	2005	10	US-09-735-367B-3	Sequence 3, Appli
12	267	3.8	2062	10	US-09-735-367B-2	Sequence 2, Appli
13	264.5	3.8	1143	12	US-10-124-557-58	Sequence 58, Appl
14	264.5	3.8	1313	12	US-10-124-557-142	Sequence 142, App
15	264.5	3.8	1314	12	US-10-124-557-50	Sequence 50, Appl
16	264.5	3.8	1354	12	US-10-124-557-48	Sequence 48, Appl
17	264	3.8	341	12	US-10-124-557-14	Sequence 14, Appl
18	264	3.8	1022	12	US-10-124-557-84	Sequence 84, Appl
19	264	3.8	1028	12	US-10-124-557-74	Sequence 74, Appl
20	264	3.8	1270	12	US-10-124-557-44	Sequence 44, Appl
21	264	3.8	1311	12	US-10-124-557-42	Sequence 42, Appl
22	264	3.8	1320	12	US-10-124-557-46	Sequence 46, Appl
23	264	3.8	1320	12	US-10-124-557-60	Sequence 60, Appl
24	264	3.8	1361	12	US-10-124-557-40	Sequence 40, Appl
25	246	3.5	1244	9	US-10-029-180-110	Sequence 110, App
26	241.5	3.5	507	9	US-10-078-547-24	Sequence 24, Appl
27	241.5	3.5	2783	10	US-09-816-669A-14	Sequence 14, Appl
28	241	3.4	1070	10	US-09-735-367B-6	Sequence 6, Appli
29	239	3.4	1215	9	US-10-029-180-34	Sequence 34, Appl
30	234.5	3.4	4019	9	US-09-854-133-425	Sequence 425, App
31	234.5	3.4	4019	10	US-09-778-977-425	Sequence 425, App
32	232	3.3	503	9	US-10-078-547-2	Sequence 2, Appli
33	232.5	3.3	881	10	US-09-816-860A-2	Sequence 2, Appli
34	230	3.3	1567	10	US-09-835-232-2	Sequence 2, Appli
35	227	3.2	802	10	US-09-823-240-2	Sequence 2, Appli
36	223	3.2	1262	10	US-09-815-242-14009	Sequence 14009, A
37	220.5	3.2	731	9	US-10-086-464-17	Sequence 17, Appl
38	219	3.1	945	8	US-09-965-272-2	Sequence 2, Appli
39	218	3.1	945	12	US-10-121-882-2	Sequence 2, Appli
40	216.5	3.1	1142	10	US-09-899-651-2	Sequence 2, Appli
41	216	3.1	2442	12	US-10-109-986-10	Sequence 10, Appl
42	215.5	3.1	1142	9	US-10-085-108-7	Sequence 7, Appli
43	212	3.0	550	9	US-09-976-740-47	Sequence 47, Appl
44	212	3.0	550	12	US-10-023-529-47	Sequence 47, Appl
45	212	3.0	550	10	US-10-023-523-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-09-803-126-1
; Sequence 1, Application US/09803126
; Patent No. US20020090190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R
; APPLICANT: Peng, Gary G
; APPLICANT: Pubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein: Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1337
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: mouse myosin related protein (MRP)
US-09-803-126-1

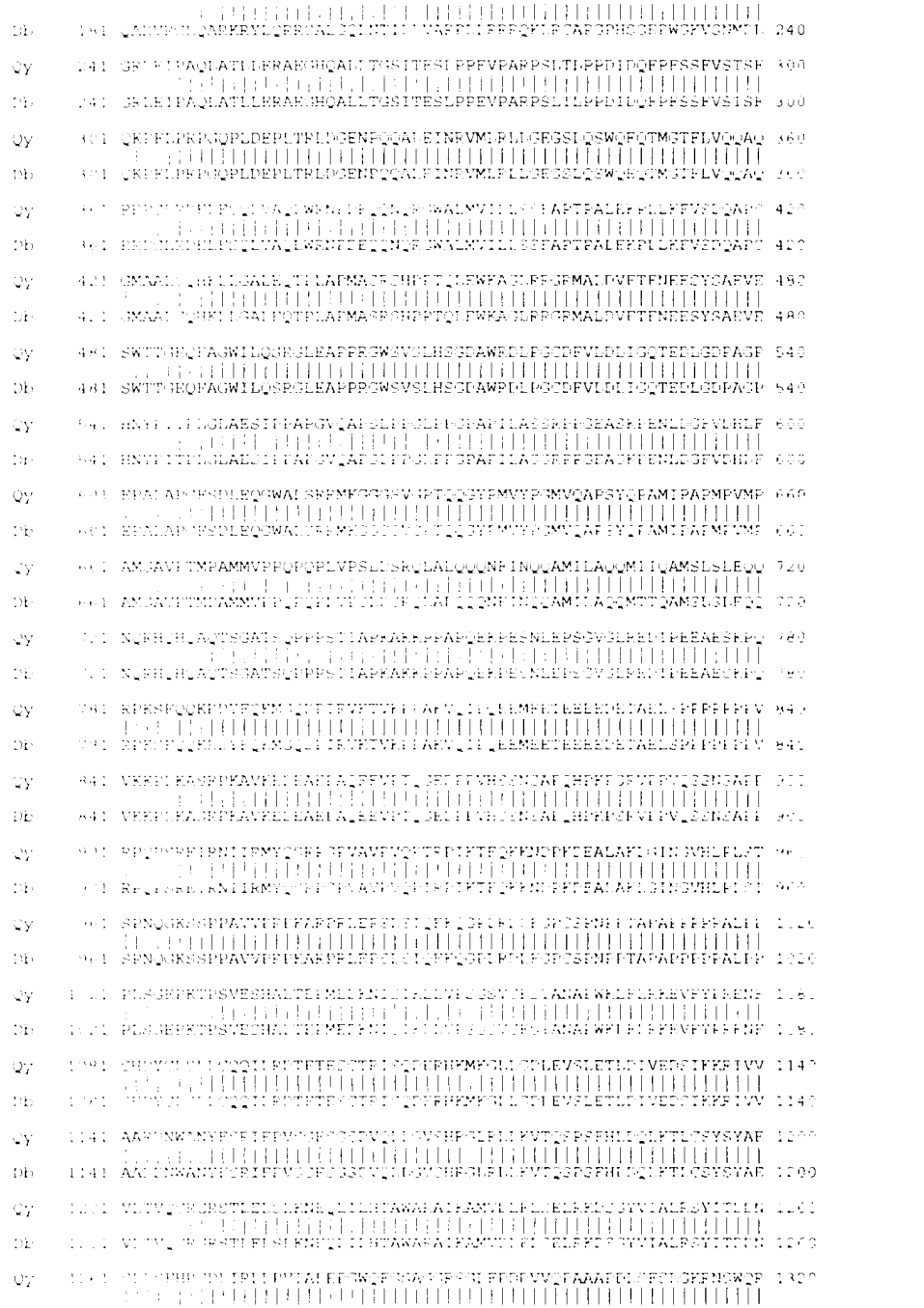
Query Match 100.0%; Score 6991; DB 10; Length 1337;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSLEQEAEFQATEQNKFRTLASFFSQTLGSLLAPLGRFGHVTVIHCLNPTPGKIPGLLD 60
Db 1 MGSLEQEAEPQAGTEQNKFRTLASFFSQTLGSLLAPLGRFGHVTVIHCLNPTPGKIPGLLD 60

Qy 61 VGHVAEQLRQAGILEIIGTRSTHFPVPVSFQVFLAPFHALGSGPQKAASDQFRCGAILSE 120
Db 61 VGHVAEQLRQAGILEIIGTRSTHFPVPVSFQVFLAPFHALGSGPQKAASDQFRCGAILSE 120

Qy 121 VLGAESPLYHLGVTQVLIQEQGWQQLPQLWAQFRSQALLTLHFGLRACITRQRLKLLFRM 180
Db 121 VLGAESPLYHLGVTQVLIQEQGWQQLPQLWAQFRSQALLTLHFGLRACITRQRLKLLFRM 180

Qy 181 QARVFGLQAFFYLQFFAIGSNLNTILVARPLLFFFIFFSFGNHFVVIMDL 240



Db 1261 SLLSPHRGDLIRLLPVTALPPGWQPGSAVRNGLPPPGVVLIAAAIIIPPTLPENGW 1320

Qy 1321 KSKLGPAQEVPRTEEVR 1337
||||||||||||||||
Db 1321 KSKLGPAQEVPRTEEVR 1337

RESULT 2
US-09-803-126-12
; Sequence 12, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Extracellular Regulated [illegible] Myosin Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-009310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: partial amino acid sequence of extracellular myosin
; OTHER INFORMATION: related protein (mMNP)
US-09-803-126-12

Query Match 29.5%; Score 2141; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.2e-104;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 914 MYQLRKQLVAVEVLQLPLPIKIFLKKENQPPEALALKLNNPVPLPLTTINQDFSPTAVV 973
Db 1 MYQLRKQLVAVEVLQLPLPIKIFLKKENQPPEALALKLNNPVPLPLTTINQDFSPTAVV 60

Qy 974 IPEPAPAPPLLELSLIQPAQPLELLEIPPNITAPAPPPPAPLPPPLSQRIPTPQVEE 1033
Db 61 IPEPAPAPPLLELSLIQPAQPLELLEIPPNITAPAPPPPAPLPPPLSQRIPTPQVEE 120

Qy 1034 HALTEIMELFNICTPLLVLCSSVTPCYAFAIWFLPLPPPIVIFPPNRQHPYYLSLDVQVI 1093
Db 121 HALTEIMELFNICTPLLVLCSSVTPCYAFAIWFLPLPPPIVIFPPNRQHPYYLSLDVQVI 180

Qy 1094 IPDIPPPYLPKISQPPHPMPLGLQVSPITLDGVPVPKFGLLDTAAPITVWANVPSPI 1153
Db 181 LPLIPPPLLPKISQPPHPMPLGLQLSPITLDGVPVPKFGLLDTAAPITVWANVPSPI 240

Qy 1154 PPVLSEQSLVLLSVHPSGLLLPVLVPSHHLQARTLCYSYVAFVLTVQCPTGTLLE 1213
Db 241 PPVLSEQSLVLLSVHPSGLLLPVLVPSHHLQARTLCYSYVAFVLTVQCPTGTLLE 300

Qy 1214 LIPNLTAWAPAIPAMVLFLLSLPALVLALLYTITNSLLSPHRGDLIRL 1263
Db 301 LIPNLTAWAPAIPAMVLFLLSLPALVLALLYTITNSLLSPHRGDLIRL 350

Qy 1264 LPVTALPPGWQPGSAVRNGLPPPGVVLIAAAIIIPPTLPENGWLP 1307
Db 351 LPVTALPPGWQPGSAVRNGLPPPGVVLIAAAIIIPPTLPENGWLP 394

RESULT 3
US-09-803-126-4
; Sequence 4, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.

; APPLICANT: Rubanyi, Gabor M
; TITLE OF INVENTION: Estrogen Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 1 (hMRP1)
US-09-803-126-4

Query Match 22.1%; Score 1543.5; DB 10; Length 437;
Best Local Similarity 71.3%; Pred. No. 6.2e-74;
Matches 308; Conservative 94; Mismatches 77; Indels 13; Gaps 2;

Qy 914 MYQSPPGPVAVPVQPTPPIKTPIPPNDPPTEALAPLGINGVH-IPLSTSPNQGKSSPPAV 972
Db 1 MYQSPPGPVEVPVQPSPPIPAPLPPIDPPIEALAPLGINGAHSSPPMLSPPPGKSPPPAV 60

Qy 973 VPPPPAPPPLEPPLSIQEPQGPLPDLPGPGSPNPPTAPAPPPPPALPPPLSGPPPTPSVE 1032
Db 61 APPPPAPLQLGPSSSIKPPQGPLLDLPG---QPLPIAHTPPPPPAPPLPLPPDPPGTLSAE 117

Qy 1033 SHALTEPMEDPNISTKLLVPSGSVCPSYANAPWPLPLPPEVPYPPPENPSHPYYLPLLCQQ 1092
Db 118 PPTLTIPVEPQGVSIQLLAPSGSVCPSYTSTPWPLPLPPEVPYPPPENPSHPYYLPLLCQQ 177

Qy 1093 ILPDTPTESGTPISLQEPHPMPGLLGLEVSLETLPIVPSIPPPIVVAAPLNWANYPSR 1152
Db 178 ILPDTPSESGTPISLQEPPPMPGLLGLEVSLETLPIVVAAPLNWANYPSR 237

Qy 1153 IPPVSGESGSDVQLLGVSHRGLPLLPVTQSPSPHLDQLPTLCSYSYAEVLTVQPGPSTL 1212
Db 238 PPPVSGESGSDVQLLAVSHRGLPLLPVTQGPGLPPDQLPILCSYSPAEVLGVEPGGSTL 297

Qy 1213 PISLPNEQLILHTAWAPAIPAMVDLPLPELPPDSGYVIALPSYITTDNSLLSFHPGPLIP 1272
Db 298 PISLPSEQIVLHTAPAPAIPALVPLPLPELPPDSGYVIALPSYITTDNSLLSFHPGPLIP 357

Qy 1273 LIPVTALEPGWQPGSAGGPSGLFPADIVQPAAAPDLSPSLGKPNSWQP---------KSK 1323
Db 358 LIPVATLEPGWQPGSAGGPSGLFPADIVQPAAAPDPSPSKPQPSGWHKGQLSNGRPSLAP 417

Qy 1324 LGPAQPVPKTPP 1335
Db 418 WDPASPVPPMGP 429

RESULT 4
US-09-803-126-13
; Sequence 13, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan P.
; APPLICANT: Deng, Gary G
; APPLICANT: Rubanyi, Gabor M
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence for human myosin
; OTHER INFORMATION: related protein variant 1 (hMRP1)
US-09-803-126-13

Query Match 22.0%; Score 1537; DB 10; Length 405;
Best Local Similarity 74.0%; Pred. No. 1.2e-73;
Matches 302; Conservative 32; Mismatches 70; Indels 4; Gaps 2;

Qy 914 MYQSPPGPVAVPVQPTPPIKTPIPPNDPPTEALAPLGINGVH-IPLSTSPNQGKSSPPAV 972
Db 1 MYQSPPGPVEVPVQPSPPIPAPLPPIDPPIEALAPLSINGAHSSPPMLSPPPGKSPPPAV 60

Qy 973 VPPPPAPPPLEPPLSIQEPQGPLPDLPGPGSPNPPTAPAPPPPPALPPPLSGPPPTPSVE 1032
Db 61 APPPPAPLQLGPSSSIKPPQGPLLDLPG---QPLPIAHTPPPPPAPPLPLPPDPPGTLSAE 117

Qy 1033 SHALTEPMEDPNISTKLLVPSGSVCPSYANAPWPLPLPPEVPYPPPENPSHPYYLPLLCQQ 1092
Db 118 PPTLTIPVEPQGVSIQLLAPSGSVCPSYTSTPWPLPLPPEVPYPPPENPSHPYYLPLLCQQ 177

Qy 1093 ILPDTPTESGTPISLQEPHPMPGLLGLEVSLETLPIVPSIPPPIVVAAPLNWANYPSR 1152
Db 178 ILPDTPSESGTPISLQEPPPMPGLLGLEVSLETLPIVVAAPLNWANYPSR 237

Qy 1153 IPPVSGESGSDVQLLGVSHRGLPLLPVTQSPSPHLDQLPTLCSYSYAEVLTVQPGPSTL 1212
Db 238 PPPVSGESGSDVQLLAVSHRGLPLLPVTQGPGLPPDQLPILCSYSPAEVLGVEPGGSTL 297

Qy 1213 PISLPNEQLILHTAWAPAIPAMVDLPLPELPPDSGYVIALPSYITTDNSLLSFHPGPLIP 1272
Db 298 PISLPSEQIVLHTAPAPAIPALVPLPLPELPPDSGYVIALPSYITTDNSLLSFHPGPLIP 357

Qy 1273 LIPVTALEPGWQPGSAGGPSGLFPADIVQPAAAPDLSPSLGKPNSWQP---------KSK 1323
Db 358 LIPVATLEPGWQPGSAGGPSGLFPADIVQPAAAPDPSPSKPQPSGWHK 405

RESULT 5
US-09-803-126-6
; Sequence 6, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan P.
; APPLICANT: Deng, Gary G
; APPLICANT: Rubanyi, Gabor M
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; TITLE OF INVENTION: Protein, Compositions and Methods of Use
; FILE REFERENCE: 015303-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Partial amino acid sequence of human myosin
; OTHER INFORMATION: related protein variant 2 (hMRP2)
US-09-803-126-6

Query Match 22.0%; Score 1537; DB 10; Length 786;
Best Local Similarity 74.0%; Pred. No. 2.2e-73;
Matches 302; Conservative 32; Mismatches 70; Indels 4; Gaps 2;

Qy 914 MYQSPPGPVAVPVQPTPPIKTPIPPNDPPTEALAPLGINGVH-IPLSTSPNQGKSSPPAV 972

```
Db          1 MYQRRPRPVPVPVQPSRRPEAFLRRILEDEALAKLDINGAHSSPPMLSPSPGRGPPPAV 60

Qy        973 VPPPKAPPPLEPSLSIQRKQGPLRPLEGPGSPNPPTAPAPPPPPALPPPLSGEPKTPSVE 1032

Db         61 APPPPAPPLQLDRSSPIPPPQQPLIPLEG---QPLPIAHTPPPPAPPLPIPEDPGTLSAE 117

Qy       1033 SHALTEPMEDKNLSTHLLVPEQSVIFSYANAPWHLFLKKEVFYPRENFSHPYCLSLLCQQ 1092

Db        118 RRLLTIIVETIQVLTLLATGSSVCFSYTGTPWKLFLKKEVFYPRENFSHPYYLRLLCEQ 177

Qy       1093 ILRDTFSESCIRISQDERRKMKDLLGGLEVSLETLDIVEDSIKKFIVVAAPDNWANYFSE 1152

Db        178 ILRDTFSESCIRISQNERRKMKDLLGGLEVDLDSLTTTEDSVKKRIVVAARDNWANYFSR 237

Qy       1153 IFPVVSEEDSDDVQLLEVSHRGLRLLKVTQGPGSHLGQLKTLCSYGYAEVLTVQTPGRSTL 1212

Db        238 FFPVSSESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAEVLGVECRGGSTL 297

Qy       1213 ELSLKNEQLILHTAWAPAIKAMVDLFLSELKKDSGYVIALRSYITDDNSLLSFHRGDLIF 1272

Db        298 ELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDNCSLLSFHRGDLIF 357

Qy       1273 LLPVTALEPGWQFGSAGGRSGLFPDDVVQPAAAPDLSFSLGKRNSWQR 1320

Db        358 LLPVAALEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHK 405
```

RESULT 6
US-09-764-868-758
; Sequence 758, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ032
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 758
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-758

Query Match 17.2%; Score 1003.5; DB 9; Length 333;
Best Local Similarity 72.8%; Pred. No. 2.7e-56;
Matches 236; Conservative 29; Mismatches 55; Indels 4; Gaps 1;

```
Qy       1021 PLSCHPETPDVESHALTEPMEDKNISTHLLVPGSSVIFSYANAPWHLFLKKEVFYPRENF 1080

Db          2 PLPEDPGTLSAERRCLTQIVETIQVETLLALSSVCFSYTGTPWKLFLKKEVFYPRENF 61

Qy       1081 SHPYCLSLLCQQILRDTFTESCTRISQDERRKMKGLLGDLEVSLETLDIVEDSIKKRIVV 1140

Db         62 SHPYYLRLLCEQILRDTFSESCIRISQNERRKMKDLLGGLEVDLDSLTTTEDSVKKRIVV 121

Qy       1141 AARDNWANYFSRIFPVSGESGSDVQLLCVSHRGLRLLKVTQGPSFHLDQLKTLCSYSYAE 1200

Db        122 AARDNWANYFSRFFPVSGESGSDVQLLAVSHRGLRLLKVTQGPGLRPDQLKILCSYSFAE 181

Qy       1201 VLTVQCRGPSTLELSLKNEQLILHTAWAPAIKAMVDLFLSELKKDSGYVIALRSYITDDN 1260

Db        182 VLGVECRGGSTLELSLKSEQLVLHTARARAIEALVELFLNELKKDSGYVIALRSYITDNC 241

Qy       1261 SLLSFHRGDLIRLLPVIALEPGWQFGSAGGRSGLFPDDVVQPAAAPDLSFSLGKRNSWQR 1320

Db        242 SLLSFHRGDLIKLLPVATLEPGWQFGSAGGRSGLFPADIVQPAAAPDFSFSKEQRSGWHK 301

Qy       1321 ---KSKLGPAQEVRKTEE 1335

Db        302 GQLSNGEPGLARWDRASEVRKMGF 325
```

RESULT 7
US-10-124-557-104
; Sequence 104, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match 3.8%; Score 267.5; DB 11; Length 1140;
Best Local Similarity 21.6%; Pred. No. 9.4e-07;
Matches 183; Conservative 79; Mismatches 350; Indels 231; Gaps 47;

```
Qy        282 TLPPDIPQPPSSPVSTSPQKPPLPRPCQPLPRPLTE  LPSENPQQALRINRVMLRLLD 339

Db        111 TSPPSSAKAPPPSGASQII-KSTIKPSKKPPNKKIKAVIKPRITTKKPVCENQ       164

Qy        340 EGSLQSWQEQTMGTFLVQQAQRRPGLRDELFSQLVAQLWRNIDEQQNQRGWALMVILLSS 399

Db        165 EESSSSSSSSSIIIWRIRLLRKLAAEKLLRKLEVE      IKEENE          TE 209

Qy        400 FAPTPALFKPLLKFVSDQAPSGM----  ---AALCQH KILQALPQTPLAPMAQR S 440

Db        210 KKPII    KTP   TTKASSSLLNIPIEVPIIIKK.IIKKP ...PIIIARPINPKPS 2??

Qy        447 HPPTQLEWKAGLRRGRMALDVFTFNEESYSAEVESWTTGEQPAGWIIQSRGLEAPPRGWS 506
```